

Result No.	Score	Query Match	Length	DB ID	Description
1	1173	100.0	1173	21	AAZ27450
2	226.4	19.3	1098	21	AAZ27462
3	200	17.1	1121	21	AAZ27460
4	195.8	16.7	796	21	AAZ27461
5	192.4	16.4	942	21	AAZ27456
6	179	15.3	3395	19	AAV61023
7	179	15.3	3395	19	AAZ45839
8	179	15.3	3395	19	AAV61024
9	179	15.3	7560	21	AAZ45840
10	177	15.1	627	19	AAV61022
11	177	15.1	627	21	AAZ45838

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 25, 2001, 09:21:45 ; Search time 9.89 Seconds
(without alignments)
962.894 Million cell updates/sec

Title: US-09-435-054-2

Perfect score: 1514

Sequence: 1 MDSSFLPAAGNCSAAG.....ASHTPQSGGLEHPPHPPAYK 278

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	391	25.8	179	1	CBFA_MAIZE
2	325.5	21.5	209	1	CBFA_PETMA
3	324.5	21.4	207	1	CBFA_HUMAN
4	324.5	21.4	207	1	CBFA_MOUSE
5	322	21.3	151	1	CBFA_CHICK
6	295.5	19.5	144	1	HAP3_YEAST
7	274	18.1	116	1	HAP3_SCHPO
8	265	17.5	205	1	HAP3_KLULA
9	232.5	15.4	122	1	CBFA_XENLA
10	195	12.9	159	1	DR1_ARATH
11	179	11.8	369	1	MAF2_RAT
12	176.5	11.7	370	1	MAF2_MOUSE
13	171	11.3	590	1	HMF2_DROME
14	170	11.2	369	1	TNAF_AVT54
15	161	10.6	176	1	TBAF_HUMAN
16	149	9.8	428	1	FXB2_MOUSE
17	147	9.7	1185	1	DRPL_HUMAN
18	139	9.2	421	1	BR3A_MOUSE
19	139	9.2	423	1	BR3A_HUMAN
20	135.5	8.9	1183	1	DRPL_RAT
21	134	8.9	549	1	DSX_DROME
22	132.5	8.8	497	1	FXD2_HUMAN
23	132.5	8.8	512	1	ANX7_XENLA
24	129.5	8.6	200	1	GRP_HORVU
25	129	8.5	440	1	FXGA_CHICK
26	129	8.5	553	1	FXC1_MOUSE
27	127.5	8.4	911	1	ANDR_MOUSE
28	124.5	8.2	895	1	ANDR_PANTR
29	124.5	8.2	919	1	ANDR_MACFA
30	123.5	8.2	269	1	MRG1_MOUSE
31	123.5	8.2	465	1	FXD1_HUMAN
32	123.5	8.2	895	1	ANDR_PAPHA
33	123	8.1	402	1	POXM_DROME

ALIGNMENTS

RESULT 1

ID CBFA_MAIZE

AC P25209; STANDARD; PRT; 179 AA.

DT 01-MAY-1992 (Rel. 22, Created)

DT 01-MAY-1992 (Rel. 22, Last sequence update)

DT 01-OCT-1996 (Rel. 34, Last annotation update)

DE CCAAT-BINDING TRANSCRIPTION FACTOR SUBUNIT A (CBF-A) (NF-Y PROTEIN

CHAIN B) (NF-YB) (CAAT-BOX DNA BINDING PROTEIN SUBUNIT B).

DE NFY2.

OS Zea mays (Maize).

OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;

OC Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae;

OC Andropogoneae; Zea.

OX NCBI_TaxID=4577;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=92195809; PubMed=1549471;

RA Li X.-Y., Mantovani R., Hooft van Huijsduijnen R., Andre I.,

RA Benoist C., Mathis D.;

RT "Evolutionary variation of the CCAAT-binding transcription factor

NF-Y.";

RL Nucleic Acids Res. 20:1087-1091(1992)

CC -FUNCTION: STIMULATES THE TRANSCRIPTION OF VARIOUS GENES BY

CC RECOGNIZING AND BINDING TO A CCAAT MOTIF IN PROMOTERS, FOR

CC EXAMPLE IN TYPE I COLLAGEN, ALBUMIN AND BETA-ACTIN GENES.

CC -SUBUNIT: HETEROMERIC TRANSCRIPTION FACTOR COMPOSED OF TWO

CC COMPONENTS, A AND B, THAT ARE BOTH NEEDED FOR DNA BINDING.

CC -SUBCELLULAR LOCATION: NUCLEAR.

CC -DOMAIN: CAN BE DIVIDED INTO THREE DOMAINS: THE WEAKLY CONSERVED A

CC SUBUNIT, THE HIGHLY CONSERVED B DOMAIN THOUGHT TO BE INVOLVED IN

CC DOMINANT INTERACTION AND DNA BINDING, AND THE GLU-RICH C DOMAIN.

CC -SIMILARITY: BELONGS TO THE CBF-A SUBUNIT FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch)

CC -----

CC EMBL: X59714; CAA42234.1; --

CC HSSP: P19267; IBFM.

CC MaizeDB: 69282; --

CC InterPro: IPR000947; --

CC Pfam: PF00808; CBFDF_NFYB_HMF: 1.

CC PRINTS: PR00615; CCAATSUBUNTA.

CC PROSITE: PS00685; CBF_A_NFYB, 1

CC Transcription regulation; DNA-binding; Activator; Nuclear protein.

CC DOMAIN 1 29 A DOMAIN.

CC DOMAIN 30 119 B DOMAIN.

CC DOMAIN 120 179 C DOMAIN.

CC DNA_BIND 36 42 BY SIMILARITY.

CC SEQUENCE 179 AA; 18995 MW; 30621316CE469454 CRC64;

34 121.5 8.0 641 1 EBN1_EBV
35 120 7.9 456 1 FXD1_MOUSE
36 120 7.9 884 1 ANDR_EULFC
37 119.5 7.9 533 1 ZIC2_HUMAN
38 118 7.8 1111 1 MYSB_DICDI
39 117.5 7.8 663 1 DUS8_MOUSE
40 117.5 7.8 1319 1 MNL_HUMAN
41 117 7.7 485 1 BIND_STRFN
42 116 7.7 1533 1 PUM_DROME
43 116 7.7 1806 1 CAIB_HUMAN
44 115.5 7.6 345 1 SOD_DROME
45 115.5 7.6 355 1 OTX1_RAT

P03211 epstein-bar
Q61345 mus musculu
O97776 eulemur ful
O95409 homo sapien
P34092 dictyosteli
O09112 mus musculu
O10571 homo sapien
P23118 strongyloce
P25822 drosophila
P2107 homo sapien
Q08473 drosophila
Q63410 rattus norv

Query Match 25.8%; Score 391; DB 1; Length 179;
 Best Local Similarity 44.6%; Pred. No. 3.4e-19;
 Matches 82; Conservative 31; Mismatches 45; Indels 26; Gaps 5;

QY 8 PAAGA---ENGSAAGANNNGAAQAAPAIREDRLMPIANVIRIRVLPAAHAKISDD 64
 DB 8 PGGGGSHESGPRGGGGG-----SVREQDRFLPIANIRIMKKAIPANGKIAKD 58
 QY 65 AKETQECVSEYISFITGEANERQREKRTITADVLWMSRLGFDYVPELGAHLRY 124
 DB 59 AKETQECVSEYISFITGEANERQREKRTITADVLWMSRLGFDYVPELGAHLRY 118
 QY 125 REFEGDARGVGLVGAAPSRGDD-----HHPHSMSPAAMLKSRGPVS-GAAMLPHHHH 176
 DB 119 REMEGDSK-----LTAKSSDGSITKDALGHVGASSAAEGCGQCGAYNQMGYQPOYH 172
 QY 177 HHDM 180
 DB 173 NGDI 176

RESULT 2
 CBFA_PETMA STANDARD; PRT; 209 AA.
 ID AC P25210;
 DT 01-MAY-1992 (Rel. 22, Created)
 DT 01-MAY-1992 (Rel. 22, Last sequence update)
 DE CCAAT-BINDING TRANSCRIPTION FACTOR SUBUNIT A (CBF-A) (NF-Y PROTEIN CHAIN B)
 DE CHAIN B) (NF-YB) (CAAT-BOX DNA BINDING PROTEIN SUBUNIT B).
 OS Petromyzon marinus (Sea lamprey).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Hyperoartia;
 OC Petromyzoniformes; Petromyzontidae; Petromyzon.
 OX NCBI_TaxID=7757;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92195809; PubMed=1549471;
 RA Li X.-Y., Mantovani R., Hooft van Hujsduijn R., Andre I.,
 RA Benoist C., Mathis D.;
 RT "Evolutionary variation of the CCAAT-binding transcription factor
 NF-Y.";
 RL Nucleic Acids Res. 20:1087-1091(1992).
 CC -!- FUNCTION: STIMULATES THE TRANSCRIPTION OF VARIOUS GENES BY
 CC RECOGNIZING AND BINDING TO A CCAAT MOTIF IN PROMOTERS, FOR
 CC EXAMPLE IN TYPE I COLLAGEN, ALBUMIN AND BETA-ACTIN GENES.
 CC -!- SUBUNIT: HETEROMERIC TRANSCRIPTION FACTOR COMPOSED OF TWO
 CC COMPONENTS, A AND B, THAT ARE BOTH NEEDED FOR DNA BINDING.
 CC -!- SUBCELLULAR LOCATION: NUCLEAR.
 CC -!- DOMAIN: CAN BE DIVIDED INTO THREE DOMAINS: THE WEAKLY CONSERVED A
 CC DOMAIN, THE HIGHLY CONSERVED B DOMAIN THOUGHT TO BE INVOLVED IN
 CC SUBUNIT INTERACTION AND DNA BINDING, AND THE GLU-RICH C DOMAIN.
 CC -!- SIMILARITY: BELONGS TO THE CBF-A SUBUNIT FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; X59712; CAA42232.1;
 CC HSP; P19267; Ibfm.
 CC InterPro: IPR000947;
 CC Pfam: PF00808; CBF2_NFYB_HME; 1.
 CC PRINTS; PS00615; CCAATSUBUNTA.
 CC PROSITE; PS00685; CBFA_NFYB; 1.
 CC Transcription regulation; DNA-binding; Activator; Nuclear protein.
 CC DOMAIN 1 53 A DOMAIN.
 CC FT DOMAIN 54 143 B DOMAIN.
 CC FT DOMAIN 144 209 C DOMAIN.
 CC FT DNA_BIND 60 66 BY SIMILARITY.

SQ SEQUENCE 209 AA; 22676 MW; E89B0F0289882350 CRC64;
 Query Match 21.5%; Score 325.5; DB 1; Length 209;
 Best Local Similarity 53.2%; Pred. No. 7e-15;
 Matches 66; Conservative 18; Mismatches 37; Indels 3; Gaps 1;

QY 11 GAENGSAAGANNNGAAQAAPAIREDRLMPIANVIRIRVLPAAHAKISDDAKETIQ 70
 DB 32 GDAEGSLASGDHDESCGSK---DPYREQDIYLPANVARIMKTSIPSSGKIADAKECVQ 88
 QY 71 ECVSEYISFITGEANERQREKRTITADVLWMSRLGFDYVPELGAHLRYREFEGD 130
 DB 89 ECVSEYISFITGEANERQREKRTITADVLWMSRLGFDYVPELGAHLRYREFEGD 148
 QY 131 ARGV 134
 DB 149 EKI 152

RESULT 3
 CBFA_HUMAN STANDARD; PRT; 207 AA.
 ID AC P25208;
 DT 01-MAY-1992 (Rel. 22, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DE NUCLEAR TRANSCRIPTION FACTOR Y SUBUNIT BETA (NF-Y PROTEIN CHAIN B)
 DE (NF-YB) (CAAT-BINDING TRANSCRIPTION FACTOR SUBUNIT A) (CBF-A) (CAAT-
 DE BOX DNA BINDING PROTEIN SUBUNIT B).
 GN NFYB OR HAP3.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92250488; PubMed=1577736;
 RA Li X.-Y., van Hujsduijn R., Mantovani R., Benoist C.O.,
 RA Mathis D.;
 RT "Intron-exon organization of the NF-Y genes. Tissue-specific splicing
 RT modifies an activation domain."
 RL J. Biol. Chem. 267:8984-8990(1992).
 CC [2]
 CC SEQUENCE FROM N.A.
 CC Badley Clarke J., Ting J.P.Y.;
 CC Submitted (XXX-1992) to the EMBL/GenBank/DBJ databases.
 CC [3]
 CC SEQUENCE FROM N.A.
 CC MEDLINE=92195809; PubMed=1549471;
 CC Li X.-Y., Mantovani R., Hooft van Hujsduijn R., Andre I.,
 CC Benoist C., Mathis D.;
 CC "Evolutionary variation of the CCAAT-binding transcription factor
 CC NF-Y.";
 CC Nucleic Acids Res. 20:1087-1091(1992).
 CC -!- FUNCTION: STIMULATES THE TRANSCRIPTION OF VARIOUS GENES BY
 CC RECOGNIZING AND BINDING TO A CCAAT MOTIF IN PROMOTERS, FOR
 CC EXAMPLE IN TYPE I COLLAGEN, ALBUMIN AND BETA-ACTIN GENES.
 CC -!- SUBUNIT: HETEROMERIC TRANSCRIPTION FACTOR COMPOSED OF TWO
 CC COMPONENTS, A AND B, THAT ARE BOTH NEEDED FOR DNA BINDING.
 CC -!- SUBCELLULAR LOCATION: NUCLEAR.
 CC -!- DOMAIN: CAN BE DIVIDED INTO THREE DOMAINS: THE WEAKLY CONSERVED A
 CC DOMAIN, THE HIGHLY CONSERVED B DOMAIN THOUGHT TO BE INVOLVED IN
 CC SUBUNIT INTERACTION AND DNA BINDING, AND THE GLU-RICH C DOMAIN.
 CC -!- SIMILARITY: BELONGS TO THE CBF-A SUBUNIT FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

```

CC
DR EMBL; L06145; AAA59930.1; -
DR EMBL; X59710; CAA42230.1; ALT_INIT.
DR PIR; S22817; S22817.
DR HSP; P19267; 1BFM.
DR TRANSFAC; T00154; -
DR MIM; 189904; -
DR InterPro; IPR000947; -
DR Pfam; PF00808; CBFD_NFYB_HMF; 1.
DR PRINTS; PR00615; CCAATSUBUNTA.
DR PROSITE; PS00685; CBFA_NFYB; 1.
KW Transcription regulation; DNA-binding; Activator; Nuclear protein.
FT DOMAIN 1 52 A DOMAIN.
FT DOMAIN 53 142 B DOMAIN.
FT DOMAIN 143 207 C DOMAIN.
FT DNA_BIND 59 65 BY SIMILARITY.
SQ SEQUENCE 207 AA; 22831 MW; 1ADF0AE45F3CF972 CRC64;

Query Match 21.4%; Score 324.5; DB 1; Length 207;
Best Local Similarity 56.6%; Pred. No. 8.1e-15;
Matches 64; Conservative 17; Mismatches 29; Indels 3; Gaps 1;

QY 34 AIREQRLMPANVIRMRVLPAAHAKISDDAKETIQECVSEYISFITGEANERCQREQR 93
DB 51 SFREQDYLPIANVARIMKNAIPQTGKIADKCEVCQSEFISFITSEASERCHQEK 110
QY 94 KTTAEVDLWAMSLRGFDYVEPLGAYLHRYREFEGDGVGLVPGAAPSRGG 146
DB 111 KTINGEDILFAMSTLGFDSYVEPLKLYLQKFRAMKGEKGIG---GAVTADG 160

RESULT 4
CBFA_MOUSE
ID CBFA_MOUSE STANDARD; PRT; 207 AA.
AC P22569;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE CCAAT-BINDING TRANSCRIPTION FACTOR SUBUNIT A (CBF-A) (NF-Y PROTEIN CHAIN B) (NF-YB) (CAAT-BOX DNA BINDING PROTEIN SUBUNIT B).
OS Mus musculus (Mouse), and Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxID=10090, 10116;
[1]
RN SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RP SPECIES=Mouse;
RC MEDLINE=91006004; PubMed=1698608;
RX Hooft van Huijsduijn R., Li X.-Y., Black D., Matthes H., Benoist C., Mathis D.;
RA "Co-evolution from yeast to mouse: cDNA cloning of the two NF-Y (CP-1/CBF) subunits."
RT EMBO J. 9:3119-3127(1990).
[2]
RN SEQUENCE FROM N.A.
RP SPECIES=Mouse;
RC MEDLINE=92195809; PubMed=1549471;
RX Li X.-Y., Mantovani R., Hooft van Huijsduijn R., Andre I., Benoist C., Mathis D.;
RA "Evolutionary variation of the CCAAT-binding transcription factor NF-Y."
RT Nucleic Acids Res. 20:1087-1091(1992).
[3]
RN SEQUENCE FROM N.A.
RP SPECIES=Mouse;
RC MEDLINE=92250488; PubMed=1577736;
RX Li X.-Y., Hooft van Huijsduijn R., Mantovani R., Benoist C., Mathis D.;
RA "Intron-exon organization of the NF-Y genes. Tissue-specific splicing modifies an activation domain."
RT J. Biol. Chem. 267:8984-8990(1992).

```

```

CC
DR EMBL; L06145; AAA59930.1; -
DR EMBL; X59710; CAA42230.1; ALT_INIT.
DR PIR; S22817; S22817.
DR HSP; P19267; 1BFM.
DR TRANSFAC; T00154; -
DR MIM; 189904; -
DR InterPro; IPR000947; -
DR Pfam; PF00808; CBFD_NFYB_HMF; 1.
DR PRINTS; PR00615; CCAATSUBUNTA.
DR PROSITE; PS00685; CBFA_NFYB; 1.
KW Transcription regulation; DNA-binding; Activator; Nuclear protein.
FT DOMAIN 1 52 A DOMAIN.
FT DOMAIN 53 142 B DOMAIN.
FT DOMAIN 143 207 C DOMAIN.
FT DNA_BIND 59 65 BY SIMILARITY.
SQ SEQUENCE 207 AA; 22831 MW; 1ADF0AE45F3CF972 CRC64;

Query Match 21.4%; Score 324.5; DB 1; Length 207;
Best Local Similarity 56.6%; Pred. No. 8.1e-15;
Matches 64; Conservative 17; Mismatches 29; Indels 3; Gaps 1;

QY 34 AIRQDRLMPANVIRMRVLPAAHAKISDDAKETIQECVSEYISFITGEANERCQREQ 93
DB 51 SFREQDYLPIANVARIMKNAIPQTGKIADKCEVCQSEFISFITSEASERCHQEK 110
QY 94 KTTAEVDLWMSRLGFDYVEPLGAYLHRYREFEGDGVGLVPGAAPSRGG 146
DB 111 KTINGEDILFAMSTLGFDSYVEPLKYLQKFRAMKGEKGIG---GAVTADG 160

RESULT 4
CBFA_MOUSE
ID CBFA_MOUSE STANDARD; PRT; 207 AA.
AC P22569;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE CCAAT-BINDING TRANSCRIPTION FACTOR SUBUNIT A (CBF-A) (NF-Y PROTEIN CHAIN B) (NF-YB) (CAAT-BOX DNA BINDING PROTEIN SUBUNIT B).
OS Mus musculus (Mouse), and Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxID=10090, 10116;
[1]
RN SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RP SPECIES=Mouse;
RC MEDLINE=91006004; PubMed=1698608;
RX Hooft van Huijsduijn R., Li X.-Y., Black D., Matthes H., Benoist C., Mathis D.;
RA "Co-evolution from yeast to mouse: cDNA cloning of the two NF-Y (CP-1/CBF) subunits."
RT EMBO J. 9:3119-3127(1990).
[2]
RN SEQUENCE FROM N.A.
RP SPECIES=Mouse;
RC MEDLINE=92195809; PubMed=1549471;
RX Li X.-Y., Mantovani R., Hooft van Huijsduijn R., Andre I., Benoist C., Mathis D.;
RA "Evolutionary variation of the CCAAT-binding transcription factor NF-Y."
RT Nucleic Acids Res. 20:1087-1091(1992).
[3]
RN SEQUENCE FROM N.A.
RP SPECIES=Mouse;
RC MEDLINE=92250488; PubMed=1577736;
RX Li X.-Y., Hooft van Huijsduijn R., Mantovani R., Benoist C., Mathis D.;
RA "Intron-exon organization of the NF-Y genes. Tissue-specific splicing modifies an activation domain."
RT J. Biol. Chem. 267:8984-8990(1992).

```

DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE CCAAT-BINDING TRANSCRIPTION FACTOR SUBUNIT A (CBF-A) (NF-Y PROTEIN
DE CHAIN B) (NF-YB) (CAAT-BOX DNA BINDING PROTEIN SUBUNIT B) (FRAGMENT).
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92195809; PubMed=1549471;
RA Li X.-Y., Mantovani R., Hooft van Huijsduijn R., Andre I.,
RA Benoist C., Mathis D.;
RT "Evolutionary variation of the CCAAT-binding transcription factor
RT NF-Y";
RL Nucleic Acids Res. 20:1087-1091(1992).
CC -!- FUNCTION: STIMULATES THE TRANSCRIPTION OF VARIOUS GENES BY
CC RECOGNIZING AND BINDING TO A CCAAT MOTIF IN PROMOTERS, FOR
CC EXAMPLE IN TYPE I COLLAGEN, ALBUMIN AND BETA-ACTIN GENES.
CC -!- SUBUNIT: HETEROMERIC TRANSCRIPTION FACTOR COMPOSED OF TWO
CC COMPONENTS, A AND B, THAT ARE BOTH NEEDED FOR DNA BINDING.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- DOMAIN: CAN BE DIVIDED INTO THREE DOMAINS: THE WEAKLY CONSERVED A
CC DOMAIN, THE HIGHLY CONSERVED B DOMAIN THOUGHT TO BE INVOLVED IN
CC SUBUNIT INTERACTION AND DNA BINDING, AND THE GLU-RICH C DOMAIN.
CC -!- SIMILARITY: BELONGS TO THE CBF-A SUBUNIT FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X59713; CAA42233.1; ALT_SEQ.
DR HSSP: P19267; IBFM.
DR InterPro: IPR000947; -.
DR Pfam: PF00808; CBF_D_NFYB_HMF; 1.
DR PROSITE: PS00685; CBF_A_NFYB; 1.
KW Transcription regulation; DNA-binding; Activator; Nuclear protein.
FT DOMAIN 1 50 A DOMAIN.
FT DOMAIN 51 140 B DOMAIN.
FT DOMAIN 141 >151 C DOMAIN.
FT DNA_BIND 57 63 BY SIMILARITY.
FT NON_TER 151 151
FT SEQUENCE 151 AA; 16733 MW; 3C2703043DE1A444 CRC64;
CC -----
Query Match 21.3%; Score 322; DB 1; Length 151;
Best Local Similarity 59.8%; Pred. No. 8.6e-15;
Matches 61; Conservative 16; Mismatches 25; Indels 0; Gaps 0;
QY 34 AIREQDRLMPIANVIRIMRVLPFAHAKISDDAKETQECVSEYISFITGEANERCQEQR 93
Db 49 SFREQDIYLIPIANVARIMKNAIFQTGKIADAKECVQECVSEFISITSEASERCHQERR 108
QY 94 KTTAEDVLWAMSRIGFDDYVEPLGAYLHRYREFEGDANGV 135
Db 109 KTINGEDILFAMSTGLGDSYVEPLKLYLQKFEAMKRGKIG 150
RESULT 6
HAP3_YEAST STANDARD; PRT; 144 AA.
AC P13434;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE TRANSCRIPTIONAL ACTIVATOR HAP3 (UAS2 REGULATORY PROTEIN A).
GN HAP3 OR YBL021C OR YBL0441.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=BMG1-7A;
RX MEDLINE=88174707; PubMed=2832732;
RA Hann S., Pinkham J., Wei R., Miller R., Guarente L.;
RT "The HAP3 regulatory locus of Saccharomyces cerevisiae encodes
RT divergent overlapping transcripts";
RL Mol. Cell. Biol. 8:655-663(1988).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=S288C;
RX MEDLINE=94103216; PubMed=8276800;
RA van Dyck L., Pearce D.A., Sherman F.;
RT "PIM1 encodes a mitochondrial ATP-dependent protease that is required
RT for mitochondrial function in the yeast Saccharomyces cerevisiae";
RL J. Biol. Chem. 269:238-242(1994).
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN=S288C;
RX Goffeau A., Jonniaux J.-L., Purnelle B., Skala J., de Wergifosse P.,
RA van Dyck L.;
RT Submitted (AUG-1994) to the EMBL/GenBank/DBJ databases.
RN [4]
RP MUTAGENESIS.
RX MEDLINE=94038948; PubMed=8223474;
RA Xing Y., Fikes J.D., Guarente L.;
RT "Mutations in yeast HAP2/HAP3 define a hybrid CCAAT box binding
RT domain";
RL EMBO J. 12:4647-4655(1993).
CC -!- FUNCTION: BINDS TO THE UPSTREAM ACTIVATION SITE (UAS) OF THE CYCL
CC GENE AND OTHER GENES INVOLVED IN MITOCHONDRIAL ELECTRON TRANSPORT
CC AND ACTIVATES THEIR EXPRESSION. RECOGNIZES THE SEQUENCE CCAAT.
CC -!- SUBUNIT: TWO COMPLEXES BIND CCAAT; COMPLEX I, THAT CONSISTS OF
CC HAP2/3/5 AND COMPLEX II, THAT CONSISTS OF HAP2/3/5/4.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- DOMAIN: CAN BE DIVIDED INTO THREE DOMAINS: THE WEAKLY CONSERVED A
CC DOMAIN, THE HIGHLY CONSERVED B DOMAIN THOUGHT TO BE INVOLVED IN
CC SUBUNIT INTERACTION AND DNA BINDING, AND THE GLU-RICH C DOMAIN.
CC -!- SIMILARITY: BELONGS TO THE CBF-A SUBUNIT FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M20318; AAA53538.1; -.
DR EMBL: X74544; CAA52633.1; -.
DR EMBL: Z35782; CAA84840.1; -.
DR PIR: A28123; A28123.
DR TRANSFAC: T00350; -.
DR SGD: S0000117; HAP3.
DR InterPro: IPR000947; -.
DR Pfam: PF00808; CBF_D_NFYB_HMF; 1.
DR PRINTS: PR00615; CCAATSUBUNITA.
DR PROSITE: PS00685; CBF_A_NFYB; 1.
KW DNA-binding; Transcription regulation; Activator; Nuclear protein.
FT DOMAIN 1 35 A DOMAIN.
FT DOMAIN 36 125 B DOMAIN.
FT DOMAIN 126 144 C DOMAIN.
FT DNA_BIND 42 48
FT DOMAIN 69 80
FT SEQUENCE 144 AA; 16154 MW; 7AB5027BAE420337 CRC64;
SQ
Query Match 19.5%; Score 295.5; DB 1; Length 144;
Best Local Similarity 50.4%; Pred. No. 4.3e-13;
Matches 60; Conservative 21; Mismatches 33; Indels 5; Gaps 2;

QY 8 PAAGAENSAAGGANGGAQOHAAPAIRODRMLPTIANVIRMRVLPAAHAKISDDAKE 67
 Db 13 PEDQENG--GNASSGSLQ--ISTLRQDRWLPINNARLMKNTLPPSAKVSXDAKE 67
 QY 68 TIQCEVSEYISFITGEANERCOREOKTITAEDVLNAMSRLGDDYVEPLGAYLHRYRE 126
 Db 68 CMQCEVSELFSVYSEASDRCAADKRTINGEDILSLHALGFENYAEVLKIYLAQRQ 126

RESULT 7
 ID PHP3_SCHPO STANDARD; PRT; 116 AA.
 AC P36611;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DE TRANSCRIPTIONAL ACTIVATOR PHP3.
 OS PHP3 OR SPAC23C11.08.
 GN Schizosaccharomyces pombe (Fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomycetes.
 ON NCBI_TaxID=4896;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=401;
 RX MEDLINE=94038948; PubMed=8223474;
 RA Xing Y., Fikes J.D., Guarente L.;
 RT "Mutations in yeast HAP2/HAP3 define a hybrid CCAAT box binding
 domain";
 RL EMBO J. 12:4647-4655(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972;
 RA Brown D., Churche C.M., Barrell B.G., Rajandream M.A., Wood V.;
 RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: BELONGS TO A COMPLEX THAT BINDS TO THE SEQUENCE CCAAT
 CC LOCATED UPSTREAM OF GENES INVOLVED IN MITOCHONDRIAL ELECTRON
 CC TRANSPORT.
 CC -!- SUBUNIT: BELONGS TO AN HETEROTRIMERIC CCAAT-BINDING COMPLEX (BY
 CC SIMILARITY).
 CC -!- SUBCELLULAR LOCATION: NUCLEAR.
 CC -!- DOMAIN: CAN BE DIVIDED INTO THREE DOMAINS: THE WEAKLY CONSERVED A
 CC DOMAIN, THE HIGHLY CONSERVED B DOMAIN THOUGHT TO BE INVOLVED IN
 CC SUBUNIT INTERACTION AND DNA BINDING, AND THE GLU-RICH C DOMAIN.
 CC -!- SIMILARITY: BELONGS TO THE CBF-A SUBUNIT FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).

CC -----
 CC EMBL: X75072; CAA52966.1; -;
 CC EMBL: 298559; CAB11161.1; -;
 CC PIR: S42744; S42744.
 CC HSSP: P19267; 1BPM.
 CC TRANSFAC: T01263; -;
 CC InterPro: IPR000947; -;
 CC Pfam: PF00808; CBF_NFYB_HMF; 1.
 CC PRINTS: PR00615; CCAATSUBUNTA.
 CC PROSITE: PS00685; CBF_NFYB; 1.
 CC KW DNA-binding; Transcription regulation; Activator; Nuclear protein.
 FT DOMAIN 1 6 A DOMAIN.
 FT DOMAIN 7 95 B DOMAIN.
 FT DOMAIN 96 116 C DOMAIN.
 FT DNA_BIND 12 18 BY SIMILARITY.
 FT SEQUENCE 116 AA; 12909 MW; 930DFCA773099CB5 CRC64;

Query Match

18.1%; Score 274; DB 1; Length 116;

Best Local Similarity 60.5%; Pred. No. 8.5e-12;
 Matches 52; Conservative 18; Mismatches 16; Indels 0; Gaps 0;

QY 41 LMPIANVIRMRVLPAAHAKISDDAKETIQCEVSEYISFITGEANERCOREOKTITAED 100
 Db 11 LPIANVARIMKALPENAKISKAKDCVQDCVSEFISVTGEASEQCTOERKKTITGD 70
 QY 101 VLWAMSRGLGDDYVEPLGAYLHRYRE 126
 Db 71 VLLALNTLGFENYAEVLKISLTKYRE 96

RESULT 8
 ID HAP3_KLULA STANDARD; PRT; 205 AA.
 AC P40914;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE HAP3 TRANSCRIPTIONAL ACTIVATOR.
 GN HAP3.
 OS Kluyveromyces lactis (Yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Kluyveromyces.
 ON NCBI_TaxID=28985;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CBS 2359 / IFO 1267 / NRRL Y-1140;
 RX MEDLINE=95147853; PubMed=7845362;
 RA Mulder W., Scholten I.H.J.M., de Boer R.W., Grivell L.A.;
 RT "Sequence of the HAP3 transcription factor of Kluyveromyces lactis
 predicts the presence of a novel 4-cysteine zinc-finger motif";
 RL Mol. Gen. Genet. 245:96-106(1994).
 CC -!- FUNCTION: BINDS TO THE UPSTREAM ACTIVATION SITE (UAS) OF THE CYC1
 CC GENE AND OTHER GENES INVOLVED IN MITOCHONDRIAL ELECTRON TRANSPORT
 CC AND ACTIVATES THEIR EXPRESSION. RECOGNIZES THE SEQUENCE CCAAT.
 CC -!- SUBUNIT: BELONGS TO AN HETEROTRIMERIC CCAAT-BINDING COMPLEX (BY
 CC SIMILARITY).
 CC -!- SUBCELLULAR LOCATION: NUCLEAR.
 CC -!- DOMAIN: CAN BE DIVIDED INTO THREE DOMAINS: THE WEAKLY CONSERVED A
 CC DOMAIN, THE HIGHLY CONSERVED B DOMAIN THOUGHT TO BE INVOLVED IN
 CC SUBUNIT INTERACTION AND DNA BINDING, AND THE GLU-RICH C DOMAIN.
 CC -!- SIMILARITY: BELONGS TO THE CBF-A SUBUNIT FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).

CC -----
 CC EMBL: L25779; AAC41662.1; -;
 CC PIR: S41926; S41926.
 CC InterPro: IPR000947; -;
 CC Pfam: PF00808; CBF_NFYB_HMF; 1.
 CC PRINTS: PR00615; CCAATSUBUNTA.
 CC PROSITE: PS00685; CBF_NFYB; 1.
 CC KW DNA-binding; Transcription regulation; Activator; Nuclear protein.
 FT DOMAIN 1 20 A DOMAIN.
 FT DOMAIN 21 110 B DOMAIN.
 FT DOMAIN 111 205 C DOMAIN.
 FT DNA_BIND 27 33 BY SIMILARITY.
 FT DOMAIN 54 65 IMPORTANT FOR SUBUNIT INTERACTIONS
 FT FT (BY SIMILARITY).
 FT DOMAIN 124 149 ASP/GLU-RICH (ACIDIC).
 FT DOMAIN 156 179 SER/THR-RICH.
 FT SEQUENCE 205 AA; 23122 MW; B33C7B6FBELD7E1A CRC64;

Query Match

17.5%; Score 265; DB 1; Length 205;

Best Local Similarity 54.3%; Pred. No. 5.7e-11;
 Matches 50; Conservative 17; Mismatches 25; Indels 0; Gaps 0;

QY 35 IREQDRLMPIANVIRIMRVLPAAHAKISDDAKETIOECVSEVISFITGEANERCOREQK 94
Db 20 LABQDRWLPINNVARUMKMTLPATTKVSQDAKECQECVSEFISPTSEACDRCTSGRRK 79
QY 95 TITAEVLMWSRLGDFDDVVEPLGAYLHYRE 126
Db 80 TINGEDILLSLHUALGEPENAEVVKIYLAKYQ 111
RESULT 9
ID CBFA_XENLA STANDARD; PRT; 122 AA.
AC P25211;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DE 01-NOV-1995 (Rel. 32, Last annotation update)
DE CCAAT-BINDING TRANSCRIPTION FACTOR SUBUNIT A (CBF-A) (NF-Y PROTEIN
CHAIN B) (NF-YB) (CAAT-BOX DNA BINDING PROTEIN SUBUNIT B) (FRAGMENT).
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92195809; PubMed=1549471;
RA Li X.-Y., Mantovani R., Hooft van Hujsduijn R., Andre I.,
RA Benoist C., Mathis D.;
RT "Evolutionary variation of the CCAAT-binding transcription factor
NF-Y";
RL Nucleic Acids Res. 20:1087-1091(1992).
CC -1- FUNCTION: STIMULATES THE TRANSCRIPTION OF VARIOUS GENES BY
RECOGNIZING AND BINDING TO A CCAAT MOTIF IN PROMOTERS. FOR
EXAMPLE IN TYPE I COLLAGEN, ALBUMIN AND BETA-ACTIN GENES.
CC -1- SUBUNIT: HETEROMERIC TRANSCRIPTION FACTOR COMPOSED OF TWO
COMPONENTS, A AND B, THAT ARE BOTH NEEDED FOR DNA BINDING.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- DOMAIN: CAN BE DIVIDED INTO THREE DOMAINS: THE WEAKLY CONSERVED A
DOMAIN, THE HIGHLY CONSERVED B DOMAIN THOUGHT TO BE INVOLVED IN
SUBUNIT INTERACTION AND DNA BINDING, AND THE GLU-RICH C DOMAIN.
CC -1- SIMILARITY: BELONGS TO THE CBF-A SUBUNIT FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

CC EMBL; X59709; CAA42229.1; -
DR PIR; S22819; S22819;
DR InterPro; IPR000947; -
DR PROSITE; PS00685; CBFA_NFYB; 1.
KW Transcription regulation; DNA-binding; Activator; Nuclear protein.
FT NON_TER 1 1
FT DOMAIN <1 57 B DOMAIN.
FT DOMAIN 58 122 C DOMAIN.
FT SEQUENCE 122 AA; 13498 MW; E76AD9FBC638F53A CRC64;

Query Match 15.4%; Score 232.5; DB 1; Length 122;
Best Local Similarity 50.0%; Pred. No. 4.4e-09;
Matches 49; Conservative 14; Mismatches 32; Indels 3; Gaps 2;

QY 69 IOECVSEVISITGEANERCOREQKTTITAEVLMWSRLGDFDDVVEPLGAYLHYREFE 128
Db 1 VOECVSEFISPTSEASRCHQERKKTNGEDILFAMSTIGDSYVEPLKLYLOKFEAM 60

QY 129 GDARCVGLVPGAAPSRGGD--HHPH-SMSPAAMLSKRG 163
Db 61 KGEIGIGTVTTDGLGDLAEPEFTSQIPAGLITTDG 98

RESULT 10
DRI_ARATH STANDARD; PRT; 159 AA.
ID DRI_ARATH
AC P49592;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE DRI PROTEIN HOMOLOG.
GN DRI.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=CV. COLUMBIA;
RX MEDLINE=951116318; PubMed=7816619;
RA Kuromori T., Yamamoto M.;
RT "Cloning of cDNAs from Arabidopsis thaliana that encode putative
protein phosphatase 2C and a human Drl-like protein by transformation
of a fission yeast mutant";
RL Nucleic Acids Res. 22:5296-5301(1994).
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -1- SIMILARITY: TO HUMAN PROTEIN DRI.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

CC EMBL; D38110; BAA07288.1; -
DR InterPro; IPR000947; -
DR Pfam; PF00808; CBFD_NFYB_HMF; 1.
KW Transcription; Nuclear protein.
SQ SEQUENCE 159 AA; 17942 MW; B5BCEC9E9BBF54FE CRC64;

Query Match 12.9%; Score 195; DB 1; Length 159;
Best Local Similarity 28.0%; Pred. No. 1.5e-06;
Matches 46; Conservative 32; Mismatches 60; Indels 26; Gaps 4;

QY 37 EODRLMPIANVIRIMRVLPAAHAKISDDAKETIOECVSEVISFITGEANERCOREQKTI 96
Db 11 KEDASLPKATMTKIEMLPDPVVRVARDADQLLIECCVEFINLVSSNDVCNKEDKRTI 70

QY 97 TAEDVLWMSRLGDFDDVVEPL-GAY-LHYREFEGDARGVGLVPGAAPSRGDDHHPHSWS 154
Db 71 APEHVLKALQVLGFGFYIEEVAAYEQHKYETMQDTQRXSVKNPG----- 115

QY 155 PAAMLKSRGVSQAAMLPHHHHHHHDMQHAAMYGTAVPVPPAGP 198
Db 116 -AQMTTEEAEEAQQRMF-----AEARARMNGGVSVPOPEHP 150

RESULT 11
ID MAF2_RAT STANDARD; PRT; 369 AA.
AC P54844;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE TRANSCRIPTION FACTOR MAF2 (PROTO-ONCOGENE C-MAF).
GN MAF2 OR MAF.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;

```

RN  SEQUENCE FROM N.A.
RP  STRAIN-WISTAR;
RX  MEDLINE=97190228; PubMed=9038383;
RA  Sakai M., Imaki J., Yoshida K., Ogata A., Matsushima-Hibaya Y.,
RA  Kuboki Y., Nishizawa M., Nishi S.;
RT  "Rat maf related genes: specific expression in chondrocytes, lens and
RT  spinal cord.";
RL  Oncogene 14:745-750(1997).
CC  -!- SUBCELLULAR LOCATION: NUCLEAR.
CC  -!- SIMILARITY: BELONGS TO THE BZIP FAMILY. MAF SUBFAMILY.
CC  -!- SIMILARITY: BELONGS TO THE BZIP FAMILY. MAF SUBFAMILY.
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
DR  EMBL; U56242; AAB50063.1; -.
DR  HSSP; P05412; IJUN.
KW  Proto-oncogene; Transcription regulation; DNA-binding; Activator;
KW  Nuclear protein.
FT  DOMAIN 139 146 POLY-ALA.
FT  DOMAIN 169 173 POLY-ALA.
FT  DOMAIN 180 187 POLY-HIS.
FT  DOMAIN 191 194 POLY-HIS.
FT  DOMAIN 212 220 POLY-GLY.
FT  DOMAIN 225 234 POLY-GLY.
FT  DOMAIN 284 310 BASIC MOTIF.
FT  DNA_BIND 312 333 LEUCINE-ZIPPER.
FT  DOMAIN 369 AA; 38457 MW; 288E464708DA6C7D CRC64;
SQ  SEQUENCE 369 AA; 38457 MW; 288E464708DA6C7D CRC64;

Query Match 11.8%; Score 179; DB 1; Length 369;
Best Local Similarity 30.9%; Pred. No. 3.8e-05;
Matches 60; Conservative 15; Mismatches 63; Indels 56; Gaps 10;

QY 99 EDVLWAM-----SRLGF--DDYVEPLGAYLHRYR-EFEGDARGVGLV-----PGAAP 142
DB 89 EDYWMYGTGYPQOLNPEALGFSPEDAVEALISNSHQLGGFDGYARGAQQLAAAGAGAGA 148
QY 143 SRGGDHHPHSMSPA-----AMLKSRGPVSGAAMLPHHHHHHDMOMHAAHYGGTAVPPGPP 199
DB 149 SLGGS--GEEMGPAAAVVSAVIAAAASGGA--PHYHHHH-----185
QY 200 HHGGLFMPHPOGSSHYLPYAYEPTYGGEHAMAAYTYGGAAYAPNGGSGGSGGGGSA 259
DB 186 HH-----ATGHH-----HHPTAGAPCAAGSASASAGAGGAGGPPASAGGGGGGGG 232
QY 260 SHTPQSGGLEPHPH 273
DB 233 GGT-AGAGGALHPH 245

RESULT 12
MAF2_MOUSE
ID MAF2_MOUSE STANDARD; PRT; 370 AA.
AC P54843;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE 30-MAY-2000 (Rel. 39, Last annotation update)
DE TRANSCRIPTION FACTOR MAF2 (PROTO-ONCOGENE C-MAF).
GN MAF2 OR MAF.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BALB/C; TISSUE=Cerebellum;
RX MEDLINE=95097997; PubMed=779931;

```

```

RA Kurschner C., Morgan J.I.;
RT "The maf proto-oncogene stimulates transcription from multiple sites
RT in a promoter that directs purkinje neuron-specific gene
RT expression.";
RL Mol. Cell. Biol. 15:246-254(1995).
CC -!- FUNCTION: THE C-MAF INTERACTION SITE WAS MAPPED TO THE SEQUENCE
CC 5'-[GTG[GC]N[GT]NCTCAGN-3' IN THE L7 PROMOTER. IT MAY INTERACT
CC WITH ADDITIONAL BASIC-ZIPPER PROTEINS THAT DETERMINE A SUBTYPE OF
CC MAF-RESPONSIVE ELEMENT BINDING.
CC -!- SUBUNIT: HOMO- OR HETERODIMER.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- SIMILARITY: BELONGS TO THE BZIP FAMILY. MAF SUBFAMILY.
CC -!- SIMILARITY: BELONGS TO THE BZIP FAMILY. MAF SUBFAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
DR  EMBL; S74567; AAB32820.1; -.
DR  HSSP; P05412; IJUN.
DR  TRANSFAC; T01432; -.
DR  MGD; MGI:96909; Maf.
KW  Proto-oncogene; Transcription regulation; DNA-binding; Activator;
KW  Nuclear protein.
FT  DOMAIN 139 146 POLY-ALA.
FT  DOMAIN 169 173 POLY-ALA.
FT  DOMAIN 180 187 POLY-HIS.
FT  DOMAIN 191 194 POLY-HIS.
FT  DOMAIN 207 210 POLY-SER.
FT  DOMAIN 212 220 POLY-GLY.
FT  DOMAIN 225 235 POLY-GLY.
FT  DNA_BIND 285 311 BASIC MOTIF.
FT  DOMAIN 313 334 LEUCINE-ZIPPER.
SQ  SEQUENCE 370 AA; 38655 MW; 76A9517EFF9C777C CRC64;

Query Match 11.7%; Score 176.5; DB 1; Length 370;
Best Local Similarity 33.5%; Pred. No. 5.6e-05;
Matches 67; Conservative 15; Mismatches 51; Indels 67; Gaps 14;

QY 99 EDVLWAM-----SRLGF--DDYVEPLGAYLHRYR-EFEGDARGVGLV-----PGAAP 142
DB 89 EDYWMYGTGYPQOLNPEALGFSPEDAVEALISNSHQLGGFDGYARGAQQLAAAGAGAGA 148
QY 143 SRGGDHHPHSMSPA-----AMLKSRGPVSGAAMLPHHHHHHDMOMHAAHYGGTAVPPGPP 197
DB 149 SLGGS--GEEMGPAAAVVSAVIAAAASGGA--PHYHHHH-----HA-----AG 190
QY 198 PPHGGFLMPHPOGSSHYLPYAYEP-TYGGEHAMAAYTYGGAAYAPNGG---SGDGGSGG 253
DB 191 HHHH-----PTAGAPGTAGGASSSSNGAGGA---GGGGPANTGGGGGGD 231
QY 254 GGGGSASHTFQSGGLEPHPH 273
DB 232 GGGGTA-----GAGGALHPH 246

RESULT 13
HMDP_DROME
ID HMDP_DROME STANDARD; PRT; 590 AA.
AC P07548;
DT 01-APR-1988 (Rel. 07, Created)
DT 01-APR-1988 (Rel. 07, Last sequence update)
DT 04-NOV-1995 (Rel. 32, Last annotation update)
DE HOMEOTIC DEFORMED PROTEIN.
GN DFD.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Plerygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.

```

OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Reguluski M., McGinnis N., Chadwick R., McGinnis W.;
 RT "Developmental and molecular analysis of Deformed: a homeotic gene
 RL controlling Drosophila head development.";
 RN [2]
 RP SEQUENCE OF 163-590 FROM N.A.
 RX MEDLINE=86162966; PubMed=3938363;
 RA Laughon A., Carroll S.B., Stoffer F.A., Riley P.D., Scott M.P.;
 RT "Common properties of proteins encoded by the Antennapedia complex
 RL genes of Drosophila melanogaster.";
 RN [3]
 RP "Homoio box genes of the Antennapedia and bithorax complexes of
 RT Drosophila";
 RL Cell 43:71-80(1985).
 CC -!- FUNCTION: SEQUENCE-SPECIFIC TRANSCRIPTION FACTOR WHICH IS PART OF
 CC A DEVELOPMENTAL REGULATORY SYSTEM THAT PROVIDES CELLS WITH
 CC SPECIFIC POSITIONAL IDENTITIES ON THE ANTERIOR-POSTERIOR AXIS.
 CC -!- FUNCTION: HOMEOTIC PROTEIN CONTROLLING DROSOPHILA HEAD
 CC DEVELOPMENT.
 CC -!- SUBCELLULAR LOCATION: NUCLEAR.
 CC -!- SIMILARITY: BELONGS TO THE ANTP FAMILY OF HOMEOBOX PROTEINS.
 CC "DEFORMED" SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X05136; CAA28782.1; -;
 DR EMBL; M13373; AAA28375.1; -;
 DR PIR; A26638; A26638.
 DR HSP; A24780; A24780.
 DR TRANSFAC; T00193; -;
 DR FlyBase; FBgn000439; Dfd.
 DR InterPro; IPR001356; -;
 DR InterPro; IPR001827; -;
 DR Pfam; PF00046; homeobox; 1.
 DR PRINTS; PR00024; HOMEOBOX.
 DR PRINTS; PR00025; ANTENNAPEDIA.
 DR PROSITE; PS00027; HOMEOBOX_1; 1.
 DR PROSITE; PS00032; ANTENNAPEDIA; 1.
 DR PROSITE; PS00071; HOMEOBOX_2; 1.
 KW Transcription regulation; Homeobox; DNA-binding;
 KW Developmental protein; Nuclear protein.
 FT DOMAIN 343 348 ANTP-TYPE HEXAPEPTIDE.
 FT DNA_BIND 366 425 HOMEOBOX.
 FT CONFLICT 227 227 S -> I (IN REF. 2).
 FT CONFLICT 442 442 G -> A (IN REF. 3).
 FT CONFLICT 443 443 N -> K (IN REF. 2).
 FT CONFLICT 443 443 MISSING (IN REF. 3).
 FT CONFLICT 456 456 MISSING (IN REF. 2).
 FT CONFLICT 545 545 MISSING (IN REF. 2).
 SQ SEQUENCE 590 AA; 63837 MW; 22407D2562FF24AB CRC64;

Query Match 11.38; Score 171; DB 1; Length 590;
 Best Local Similarity 30.18; Pred. No. 0.0002;
 Matches 71; Conservative 9; Mismatches 62; Indels 94; Gaps 16;
 QY 109 GFDDYVEPLGAYLHRYREF-----EGDARGVGLVPGAAPSRGGDHHPHSMSP 155
 DB 24 GLDPKFPVADDYHYHNGHYSWTASTGMSGAVGGAGVGSV-GGGGAGGTGHPHSMHP 82

QY 156 AAMLSRGPVSGAAMLPHH-----HHHMQM-----HAAMYGG-----TA 191
 DB 83 ADMVSD-----YMAHHNPHSHSHSHSHSLPHHHHSNLSAISCHHQASAGGYSSNYANA 134
 QY 192 VPP--PAGPPHHGGFLMHPH--QGSSTHLYPYAYE-----PTYGGEHMAAAYGGAA 238
 DB 135 TTPSPHSHPH-----AHPHQSLSGYVYVHHAPETISAGAVHSDPTNG-----YGPAA 180
 QY 239 YAP--GNGSGDGGSG--GGG--GGSAS-----HTPOGSGGLEH-----PH 273
 DB 181 NVPNTSNGGGGGGAGVGGGAVGSGANGYGGYGGGTANGSVGTHSGHSPH 236
 RESULT 14
 TMAF_AVISA4 STANDARD; PRT; 369 AA.
 AC P23091;
 DT 01-NOV-1991 (Rel. 20, Created)
 DT 01-NOV-1991 (Rel. 20, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE TRANSFORMING PROTEIN MAF.
 GN V-MAF.
 OS Avian musculoaponeurotic fibrosarcoma virus AS42.
 CC Viruses; Retroviral viruses; Retroviridae; Avian type C retroviruses.
 OX NCBI_TaxID=11873;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90046665; PubMed=2554284;
 RA Nishizawa M., Kataoka K., Goto N., Fujiwara K.T., Kawai S.;
 RT "v-maf, a viral oncogene that encodes a 'leucine zipper' motif";
 RL Proc. Natl. Acad. Sci. U.S.A. 86:7711-7715(1989).
 CC -!- FUNCTION: MIGHT BE A TRANSCRIPTIONAL TRANS-ACTIVATOR.
 CC -!- SUBCELLULAR LOCATION: NUCLEAR.
 CC -!- DISEASE: INDUCES MUSCULOAPONEUROTIC FIBROSARCOMA IN CHICKENS.
 CC -!- MISCELLANEOUS: THIS PROTEIN IS SYNTHESIZED AS A ENV-MAF
 CC POLYPROTEIN.
 CC -!- SIMILARITY: BELONGS TO THE BZIP FAMILY. MAF SUBFAMILY.
 CC -----

This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----

EMBL; M26769; AAA42377.1; -;
 PIR; B33975; TVFVAF.
 HSP; P05412; 1JUN.
 TRANSFAC; T01430; -;
 KW Transforming protein; Transcription regulation; DNA-binding;
 FT DOMAIN 163 167 POLY-ALA.
 FT DOMAIN 173 182 HIS-RICH.
 FT DOMAIN 184 191 POLY-GLY.
 FT DOMAIN 214 223 POLY-GLY.
 FT DOMAIN 233 240 POLY-GLY.
 FT DNA_BIND 274 300 BASIC MOTIF.
 FT DOMAIN 302 323 LEUCINE-ZIPPER.
 SQ SEQUENCE 369 AA; 38892 MW; F386B220ACE50FF6 CRC64;

Query Match 11.28; Score 170; DB 1; Length 369;
 Best Local Similarity 31.18; Pred. No. 0.00015;
 Matches 61; Conservative 9; Mismatches 52; Indels 74; Gaps 11;

QY 99 EDVLWAM-----SRLCGF--DDYVEPL-----GAYLHRYREFGDARGVGLVPG 139
 DB 89 EDYWTGTGTPQQLNPALGFSPDEAVEALINSHPLPCA-----FDYARGQQL---- 138
 QY 140 AAPSRGDDHHPHSMSPAAKLKSRGPVSGAAM--LPHHHHHMQMHAAMYGGTAVPPPPAG 197

OM of: US-09-435-054-2 to: GenEmbl.* out_format : pfs
 Date: Oct 25, 2001 2:23 PM
 About: Results were produced by the GenCore software, version 4.5,
 Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODE=frame+p2n.model -DEV=xlh
 -O=/cgn2.1/USPTO_spool/US09435054/runat_25102001_085523_6975/app_query.fasta_1.338
 -DB=GenEmbl -QFMT=fastap -SUFFIX=p2n.rge -GAPOP=12.000
 -CAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
 -GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
 -FCAPOP=6.000 -FCAPEXT=7.000 -FCAPOP=10.000 -YGAPEXT=0.500
 -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blotsum62
 -TRANS=human40.cdi -LIST=45 -DOALIGN=200 -THR_SCORE=pct
 -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
 -NORM=ext -MINLEN=0 -MAXLEN=200000000
 -USER=US09435054_cgn1_1.6559 -NCPU=6 -ICPU=3 -LONGLOG
 -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-435-054-2
 Query length: 278
 Database: GenEmbl.*
 Database sequences: 1344157
 Database length: -856060004
 Search time (sec): 1216.080000

score_list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
gb_p11:AB025628	+ 503.00	459.47	2.3e-17	80117	Arabidopsis thaliana
gb_p11:AF036684	+ 494.00	474.70	3.3e-18	826	Arabidopsis thaliana C
gb_p11:AC013482	+ 494.00	450.89	7.1e-17	82875	Arabidopsis thaliana C
gb_p12:ATFAP3B	+ 443.50	427.28	1.5e-15	874	Arabidopsis thaliana mRNA
gb_p12:ATFAP3B	+ 410.50	368.47	2.7e-12	197976	Arabidopsis thaliana
gb_p12:ATFAP3B	+ 410.50	368.25	2.8e-12	206606	Arabidopsis thaliana
gb_p12:ATFAP3B	+ 404.50	364.60	4.5e-12	141808	Arabidopsis thaliana
gb_p12:ATFAP3B	+ 404.50	363.38	5.3e-12	179686	Arabidopsis thaliana
gb_p12:ATFAP3B	+ 392.50	379.71	6.5e-13	870	Arabidopsis thaliana
gb_p12:ATFAP3B	+ 376.50	365.01	4.3e-12	832	Arabidopsis thaliana
gb_p12:ATFAP3B	+ 359.00	323.58	8.7e-10	107377	Arabidopsis thaliana
gb_p11:AC005309	+ 354.50	320.74	1.3e-09	82403	Arabidopsis thaliana
gb_p11:AC007063	+ 340.50	331.71	3.1e-10	787	Arabidopsis thaliana
gb_ov:AF041204	+ 325.50	318.32	1.7e-09	699	Arabidopsis thaliana
gb_ov:PMNFB	+ 324.50	317.97	1.8e-09	624	Arabidopsis thaliana
gb_p10:HUMFYBSU	+ 324.50	317.14	2.0e-09	734	Arabidopsis thaliana
gb_p10:RATCFB1	+ 324.50	317.10	2.0e-09	739	Arabidopsis thaliana
gb_p10:RATCFB1	+ 324.50	314.45	2.8e-09	1234	Arabidopsis thaliana
gb_p11:AC005317	+ 324.50	314.25	2.9e-09	1284	Arabidopsis thaliana
gb_p11:AC005316	+ 324.50	314.02	3.0e-09	1341	Arabidopsis thaliana
gb_ov:GGNFB	+ 322.00	314.95	2.6e-09	1343	Arabidopsis thaliana
gb_p12:AF037602	+ 313.50	303.02	1.2e-08	713	Arabidopsis thaliana
gb_p12:AF037602	+ 307.50	273.98	5.0e-07	1547	Arabidopsis thaliana
gb_p12:AF037602	+ 307.00	274.90	4.5e-07	144628	Arabidopsis thaliana
gb_p12:AF037602	+ 304.00	269.82	8.6e-07	110514	Arabidopsis thaliana
gb_p12:AF037602	+ 295.50	289.41	7.0e-08	171809	Arabidopsis thaliana
gb_p12:AF037602	+ 295.50	287.87	8.5e-08	835	Arabidopsis thaliana
gb_p12:AF037602	+ 295.50	282.70	1.6e-07	1125	Arabidopsis thaliana
gb_p12:AF037602	+ 289.00	280.34	2.2e-07	3061	Arabidopsis thaliana
gb_p12:AF037602	+ 289.00	282.80	1.6e-07	4834	Arabidopsis thaliana
gb_p12:AF037602	+ 286.00	263.03	2.1e-06	928	Arabidopsis thaliana
gb_p12:AF037602	+ 286.00	257.29	4.3e-06	75234	Arabidopsis thaliana
gb_p12:AF037602	+ 274.00	267.36	1.2e-06	266340	Arabidopsis thaliana
gb_p12:AF037602	+ 273.50	263.21	2.0e-06	1226	Arabidopsis thaliana
gb_p12:AF037602	+ 271.00	260.95	2.7e-06	2499	Arabidopsis thaliana
gb_p12:AF037602	+ 269.50	263.80	1.9e-06	1084	Arabidopsis thaliana
gb_p12:AF037602	+ 260.50	257.98	3.9e-06	658	Arabidopsis thaliana

gb_p13:ENU35341 + 246.50 236.11 6.5e-05 3618 ! U35341 Emeritella nidulans
 gb_hg18:AC084359 - 244.50 214.07 0.0011 179882 ! AC084359 Homo sapiens chr
 gb_hg19:AC023350 - 244.50 213.97 0.0011 182972 ! AC023350 Homo sapiens chr
 gb_ov:XUNFB + 241.50 242.30 2.9e-05 442 ! X59709 X.laevus mRNA for CAA
 gb_susi:CNS071BW + 241.50 238.07 5.0e-05 1003 ! AL424722 T7 end of clone XA

seq_name: gb_p11:AB025628

seq_documentation_block:

LOCUS AB025628 80117 bp DNA PLN 27-DEC-2000
 DEFINITION Arabidopsis thaliana genomic DNA, chromosome 5, p1 clone:MNJ7.
 ACCESSION AB025628 BA000015
 VERSION AB025628.1 GI:4589434
 KEYWORDS

SOURCE Arabidopsis thaliana (strain:Columbia) DNA, clone_lib:Mitsui P1

ORGANISM

Arabidopsis thaliana

Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;

Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE

1 (sites) Kaneko,T., Katoh,T., Asamizu,E., Sato,S., Nakamura,Y., Kotani,H. and Tabata,S.

AUTHORS

Structural analysis of Arabidopsis thaliana chromosome 5. XI

TITLE

Unpublished

REFERENCE

2 (bases 1 to 80117)

AUTHORS

Nakamura,Y.

TITLE

Direct Submission

JOURNAL

Submitted (02-APR-1999) Yasukazu Nakamura, Kazusa DNA Research

Institute, Department of Plant Gene Research, 1532-3, Yana,

Kisarazu, Chiba 292-0812, Japan (E-mail: ynakamura@kazusa.or.jp,

Tel:81-438-52-3935, Fax:81-438-52-3934)

Address for correspondence: kaos@kazusa.or.jp

For the latest information on annotation of this clone, please see

http://www.kazusa.or.jp/kaos/cgi-bin/sgd-graph.cgi?c=MNJ7

Genes with similarity to proteins in the databases are described in

'product' or 'note' qualifiers. Genes that have no significant

protein similarity are described as 'unknown protein'.

The software programs used to predict genes include: Grail

(Informatics Group, Oak Ridge National Laboratory,

http://compbio.ornl.gov/grail-1.3/),

GENSCAN (Chris Burge, MIT, http://ccr-081.mit.edu/GENSCAN.html),

NetGene2 (S.M. Hebsgaard, et al., CBS, Technical University of

Denmark, http://www.cbs.dtu.dk/services/NetGene2/) and

SplicePredictor (Volker Brendel, Stanford University,

http://grenlini.zool.iastate.edu/cgi-bin/sp.cgi).

Genes encoding tRNAs are predicted by tRNAscan-SE

(Sean Eddy, Washington University School of Medicine, St. Louis,

http://genome.wustl.edu/eddy/tRNAscan-SE/).

This sequence may not be the entire insert of this clone. It may be

shorter because we remove overlaps between neighboring submissions.

The 5' clone is MQL5 and the 3' clone is MGCl.

Location/Qualifiers

1. 80117

/organism="Arabidopsis thaliana"

/strain="Columbia"

/db_xref="taxon:3702"

/chromosome="5"

/clone="MNJ7"

/clone_lib="Mitsui P1"

complement(12..108)

/note="CDS is reported in Acc# AB018117"

/number=2

/evidence=not_experimental

/product="DNA-binding protein-like"

complement(375..540)

/note="CDS is reported in Acc# AB018117"

/number=1

/evidence=not_experimental

/product="DNA-binding protein-like"

Join(1723..1976,2116..2266,2344..2417,2508..2667,

CDS

3	SerSerSerPheLeuProAlaAlaGlyAlaGluAsnGlySerAlaAlaGlu	19	..
75286	GCAGCGAATTTCTGTATGAGGAGGCGAGTATCGCTCCAGAAATTCAA	75335	..
19	YGLYAlaAsn.....AsnGlyGlyAlaLaGlnGlnHisAlaAlaP	33	..
75336	CGAGCCTAAACAAACACAGTAATGGTGGTGGAGGAG.....T	75373	..
33	roAlaIleArgGluInAspArgLeuMetProIleAlaAsnValIleArg	49	..
75374	GCAGCGTGGAGGAGCAGACAGGTTCATGCCATTATGCCAACGTATACGG	75423	..
50	IleMetArgArgValLeuProAlaHisAlaLysIleSerAspAlaLys	66	..
75424	ATCATCGGAGGAGTCTTACCTGCTCACGCCAAGATCTCAGATGACTCCA	75473	..
66	sGluThrIleGlnIleCysValSerGluTyrIleSerPheIleThrGlyG	83	..
75474	GGAGACGATCCAGAGGTGTGTTTCGGAGTACATCAGCTTCATAACAGGG	75523	..
83	luAlaAsnGluArgCysGlnArgGluInArgLysThrIleThrAlaGlu	99	..
75524	AGGCTAATGAGCGGTCCAGCGGGAAACAGCGCAAGACCATCAGTCTGAG	75573	..
100	AspValLeuTrpAlaMetSerArgLeuGlyPheAspAspTyrValGluPr	116	..
75574	GAGCTCTGTGGGCAATGAGCAAGCTCGGTTTGTATGACTACATCGAACC	75623	..
116	oLeuGlyAlaTyrLeuHisArgTyrArgGluPheGluGlyAspAlaArg	133	..
75624	CCTCAGCTTGCTACTCCACCGCTACAGAGAGTGGAGAGGTAA...AGAG	75670	..
133	lyValGlyLeuValProGlyAlaAlaProSerArgGlyGlyAspHisHis	149	..
75671	GGGTAGCTGACGTGGTGGTCCGTT.....	75696	..
150	ProHisSerMetSerProAlaAlaMetLeuLys.....	160	..
75697AGTATGACCAACGCTTGTGGTCAAGAGCGCTAATGGGACCAT	75740	..
161SerArgGlyProValSerGlyAlaAlaMetLeuProH	173	..
75741	GACCGAGTATGGAGCCTACGGGCGCTGTCGCCAGGAGTTCACATGGCGAGT	75790	..
173	ishHisHisHisHisAsp	179	..
75791	ACCATTTATCGTCATCAGAAC	75810	..

seq_name: gb_pl1:AF036684
seq_documentation_block:
LOCUS AF036684 826 bp DNA PLN 02-JUL-1998
DEFINITION Arabidopsis thaliana CCAAT-box binding factor HAP3 homolog gene, complete cds.
ACCESSION AF036684
VERSION AF036684.1 GI:3282673
KEYWORDS
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
1 (bases 1 to 826)
Lotan,T., Ohto,M., Yee,K.M., West,M.A., Lo,R., Kwong,R.W., Yamagishi,K., Fischer,R.L., Goldberg,R.B. and Harada,J.J. Arabidopsis LEAFY COTYLEDON1 is sufficient to induce embryo development in vegetative cells
Cell 93 (7), 1195-1205 (1998)
98319234
2 (bases 1 to 826)
Lotan,T., Ohto,M.-A., Yee,K.M., West,M.A.L., Lo,R., Kwong,R.W., Fischer,R.L., Goldberg,R.B. and Harada,J.J.

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
REFERENCE
AUTHORS

```

450 ..... 450
182 tHisAlaAlaMetTyrGlyGlyThrAlaValProProAlaGlyProp 199
|||||.....
451 .....TATGGAGGAATGGTATT..... 468
199 roHisHISGlyGlyPheLeuMetProHisProGlnGlySerSerHisTyr 215
|||||.....
469 .....GGGTTT.....CAGGCCCATCTCATGCC 492
216 Leu.....ProTyrAlaTyrGluProThrTyrGlyGlyGluHi 228
|||||.....
493 CTACCTCTCCGGGTCTTATGGTTAT.....GGTATGTT 527
228 sAlaMetAlaAlaTyrTyrGlyGlyAlaAlaTyrAlaProGlyAsnGlyG 245
|||||.....
528 GGACCAATCATCGTTATGGAGGTGGTGGTACTACCAA...AACGGT 574
245 lySerGlyAspGlySerGlySerGlyGlyGlyGlyGlySerAlaSer 260
|||||.....
575 CGTGGGTCAAGATGAATCAGTGTGGTGGTGGCTTCGCTTCC 621

seq_name: gb_pl1:AC013482

seq_documentation_block:
LOCUS AC013482 82875 bp DNA PLN 28-JUN-2000
DEFINITION Genomic sequence for Arabidopsis thaliana BAC T26F17 from
chromosome I, complete sequence.
ACCESSION AC013482
VERSION AC013482.2 GI:6478885
KEYWORDS HTG.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 82875)
Shinn,P., Brooks,S., Buehler,E., Chao,Q., Johnson-Hopson,C.,
Khan,S., Kim,C., Altafi,H., Bei,Q., Chin,C., Chiou,J., Choi,E.,
Conn,L., Conway,A., Gonzales,A., Hansen,N., Howing,B., Koo,T.,
Lam,B., Lee,J., Lenz,C., Li,J., Liu,A., Liu,K., Liu,S.,
Mukharsky,N., Nguyen,M., Palm,C., Pham,P., Sakano,H., Schwartz,J.,
Southwick,A., Thaveri,A., Toriumi,M., Vaysberg,M., Yu,G.,
Fedorpiel,N.A., Theologis,A. and Ecker,J.R.
Genomic sequence for Arabidopsis thaliana BAC T26F17 from
chromosome I
Unpublished
2 (bases 1 to 82875)
Ecker,J.R.
Direct Submission
Submitted (13-NOV-1999) Arabidopsis thaliana Genome Center,
Department of Biology, University of Pennsylvania, 38th Street and
Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA
3 (bases 1 to 82875)
Ecker,J.R.
Direct Submission
Submitted (30-NOV-1999) Arabidopsis thaliana Genome Center,
Department of Biology, University of Pennsylvania, 38th Street and
Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA
4 (bases 1 to 82875)
Shinn,P., Brooks,S., Buehler,E., Chao,Q., Johnson-Hopson,C.,
Khan,S., Kim,C., Altafi,H., Bei,Q., Chin,C., Chiou,J., Choi,E.,
Conn,L., Conway,A., Gonzales,A., Hansen,N., Howing,B., Koo,T.,
Lam,B., Lee,J., Lenz,C., Li,J., Liu,A., Liu,K., Liu,S.,
Mukharsky,N., Nguyen,M., Palm,C., Pham,P., Sakano,H., Schwartz,J.,
Southwick,A., Thaveri,A., Toriumi,M., Vaysberg,M., Yu,G., Davis,R.,
Fedorpiel,N., Theologis,A. and Ecker,J.
Direct Submission
Submitted (09-DEC-1999) Arabidopsis thaliana Genome Center,
Department of Biology, University of Pennsylvania, 38th and
Hamilton Walk, Philadelphia, PA 19104-6018, USA
5 (bases 1 to 82875)
Cheuk,R., Shinn,P., Brooks,S., Buehler,E., Chao,Q.,
Johnson-Hopson,C., Khan,S., Kim,C., Altafi,H., Bei,Q., Chin,C.,
Chiou,J., Choi,E., Conn,L., Conway,A., Gonzales,A., Hansen,N.,
Howing,B., Koo,T., Lam,B., Lee,J., Lenz,C., Li,J., Liu,A., Liu,K., Liu,S.,
Mukharsky,N., Nguyen,M., Palm,C., Pham,P., Sakano,H., Schwartz,J.,
Southwick,A., Thaveri,A., Toriumi,M., Vaysberg,M., Yu,G., Davis,R.,
Fedorpiel,N., Theologis,A. and Ecker,J.
Direct Submission
Submitted (28-JUN-2000) Arabidopsis thaliana Genome Center,
Department of Biology, University of Pennsylvania, 38th and
Hamilton Walk, Philadelphia, PA 19104-6018, USA
On Nov 30, 1999 this sequence version replaced gi:6403470.
Location/Qualifiers
1. 82875
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"
/chromosome="1"
/clone="T26F17"
join(140..398,676..953,1145..1474)
/notes="unknown; similar to EST gb|H74506"
/codon_start=1
/evidence=not_experimental
/product="T26F17.1"
/protein_id="AAF16527.1"
/db_xref="GI:6552728"
/translation="MARNGIGSGGKNGAGOFFLATLLLLWLVVVFETAFNRTTELL
WTCGGCFQLMNMVVRWISRDPIFVNTSVSLHSLITSVSVFILLNQCLAKGLDE
MFDSBELVGTGWAYPALFCGYPAYQWDMLOYRLVSLGLPSLVHLLVLLVGLDE
LALYRNVTINYLILKILICEMISIFLHVRLKRMAGIRDSNTALVLEWLVNMTAFVFA
RCIPHILITIKLIKDAHKFGKVELPALSGMAGMNLNVGLGLDLFHAFRRERSHRR
NQENGNLHDHGE"
complement(1903..1974)
/notes="codon recognized: CAA"
/product="tRNA-Gln"
/evidence=not_experimental
complement(join(2361..2747,2834..4239,4324..4420))
/notes="hypothetical protein"
/codon_start=1
/evidence=not_experimental
/product="T26F17.2"
/protein_id="AAF16547.1"
/db_xref="GI:6552748"
/translation="MMTGTTTLLPEVMDSKDELVKQAKVAEDAVAGWEKAENEVVEL
KOKLEAADANKIVLEDRVSHLDGALKECVRLQRFDEQKNQIAAVTSTKELHSAN
TGLEKRVLEQKAEAAKSENMMLRREFLTORDELIVMTLERDLSTQAAFTASKOHL
IIRKLAKEACRRLRIKATSSLSNQSVDHSGGRRVEGSGSDSWASAFISE
LDQIKNEKGNRSQGTSTTEIDLDDFLEMERLVALPTETAKKSKGDYELSLWEK
LEKTAEKDDLEVEKCCREARLLEIEAVVGDKMELEMLKRVENAEKELKTFD
VLKDKYQESRVCFQVDTIKLEKQAEKDELDSVICKAEKRFSELEAVVGDKTIE
EDELKMEAKAELKISFDVIKQYQESRVCFQVEMKLEAMKRELKLANESKTOAES
RVTRMEAEVKERIVSDGLKEKGETPEELRRETEKMTIKREKVEPKIKQEDIATAA
KGFADCKTASGKQSLATLEEFLLDTASTPGSARSVHNKEALLGKDPHCICKTI
NGRSLEFLATQNSNNKTSPPCSSSDSTTVSLIMSSNRGSSSEKRNNGFATVTFRSRNS
VNIGI"
complement(7212..7601)
/notes="unknown; similar to EST gb|AA586241"
/codon_start=1
/evidence=not_experimental
/product="T26F17.3"
/protein_id="AAF16546.1"
/db_xref="GI:6552747"
/translation="MINLKFMFCSDSFISIKFENDGNASWEKKAIVTKVCVONNNNN
TTNCTCKKQSPITSNSTIVFHSKSSSLRWKRIGHLFHVILKSGSSSTSSCHVAPS
KVEGTVKVKQGNWVTRTLTRKSKRSKS"
complement(8023..8304)
/notes="unknown; similar to ESTs gb|T21171 and
gb|AI992884.1"
/codon_start=1
/evidence=not_experimental
/product="T26F17.4"
/protein_id="AAF16545.1"
/db_xref="GI:6552746"
/translation="MDMYKDDSPYCYFHPKEEYGVGVCPLCLNRLRLVLAKQKRSRT
KHSSSSPIISLPKIFALSSLSRLDLRRKFPSPSSDLVDSTSQECTVFL"

```

CDS

12642..13364
 /note="similar to urease accessory protein UREF sp|Q07402"
 /codon_start=1
 /evidence="not_experimental"
 /product="T26F17.5"
 /protein_id="AAFI6528.1"
 /db_xref="GI:6552729"
 /translation="MEEDERDIVMSRASSCMQWQMSQWLLDLSILPTGGFAHSFLEAA
 TOTLVSSPEDETHIHVLDNTASLLLPFYVYSAKSPDIEHKLGLNATITNOV
 SSKASMSQCSALFRITAASTFVTEPNLKMIRDAISGKNCVCFHAPRIGLVCGLLGMSQ
 ETSORAYLFVTLRVLSAATRLNIVGPMGASVMQHRRIATVETVLEKMNREAGCQ
 TSPLLDVQGGCHYLSRLFC"
 complement(join(13586..13828,13930..14008,14146..14511,
 14642..14984,15072..15171,15388..15658,15975..16084,
 16170..16313))
 /note="similar to pollen-specific protein precursor
 emb|CAA23065; similar to ESTs emb|F14023, gb|AI491058.1,
 gb|IR30156, gb|AI774044.1, gb|AI730478.1, gb|AI726799, and
 gb|AI992516.1"
 /codon_start=1
 /evidence="not_experimental"
 /product="T26F17.6"
 /protein_id="AAFI6544.1"
 /db_xref="GI:6552745"
 /translation="MEVSVNTTAMILGLFLISFVAEDPYKFFEHVHTYGNISPLK
 VAQGLILNGKFGPDIAAAYNTDNLIIINFNHLDPEFLISWGIIRNWRNSYQGVYGT
 TCPIPPKNTYALQVDQIGSFYFSLGFKAAAGFGAIRISRRPRIIPVPPAPAG
 DYTLLGDWYKTHKDLKAOLDNGKLPDPDGLINGRSGATLNIPEPKTYRLRLSN
 VGLQNSLFRIONHKMLKEVEGTHITOTPESSLDVHVQGSYVLITADOPAKDYIV
 VSSRFTSKILTTGVLYHSNAPVSGPIPEAPQLPWSFDQARAINKLAASGRPN
 PQSNHYGKIKVTPTIKLASSAGNNGKQRYAVNSAFYFTDPLKADIFKTAGVYN
 PGSPDPTGCAIYPTVSMGTDYKAEVIEFENWEDIVQTHLWDGYSFFVGMELGK
 RASRYNLDNVAISCTVOQVYPRSWTAIYVSLDNGMNLKSELWERQYLQQQFYM
 WYPTPSLNDYILPKNALLCGRATGTHHTTPGLSEGERF"
 complement(join(16811..17014,17112..17190,17407..17772,
 18036..18378,18463..18562,18737..19007,19646..19755,
 19855..19998))
 /note="similar to pectinesterase emb|CAB37498.1; similar
 to ESTs emb|F14123, gb|I75751, gb|IR30156, gb|AI730478.1,
 gb|AI774044.1, gb|AI726799.1, and gb|AI992516"
 /codon_start=1
 /evidence="not_experimental"
 /product="T26F17.7"
 /protein_id="AAFI6543.1"
 /db_xref="GI:6552744"
 /translation="MKVKSMTNTRAMITTLFLSLAFADPPYFFEHVHTYGNISPLG
 VAQGLILNGKFGPDILITNDNLIIINFNHLDPEFLISWGIIRNWRNSFQGVYGT
 MCPTIPGKNTYALQVDQIGSFYFSLGFKAAAGFGAIRISRRALIPVPPPTAD
 DYTLLGDWYKTHKDLKAOLDNGKLPDPDGLINGRSGATLNIPEPKTYRLRLSN
 VGLQNSLFRIONHKMLKEVEGRTIONLFSSLDVHVQGSYVLITADOPAKDYIV
 VSSRFTSKILTTGVLYHSNAPVSGPIPEAPQLPWSFDQARAINKLAASGRPN
 PQSNHYGKIKVTPTIKLASSAGNNGKQRYAVNSAFYFTDPLKADIFKTAGVYN
 PGSPDPTGCAIYPTVSMGTDYKAEVIEFENWEDIVQTHLWDGYSFFVGMELGK
 RASRYNLDNVAISCTVOQVYPRSWTAIYVSLDNGMNLKSELWERQYLQQQFYM
 WYPTPSLNDYILPKNALLCGRASSHR"

CDS

alignment_scores:
 Quality: 494.00 Length: 249
 Ratio: 3.146 Gaps: 9
 Percent Similarity: 63.052 Percent Identity: 46.988
 alignment_block:
 US-09-435-054-2 x AC013482/rev ..
 Align seg 1/1 to reverse of: AC013482 from: 1 to: 82875
 16 SerAlaLaGlyGlyAlaAsnGlyGlyAlaLaGlnGlnHisAlaAla 32
 74138 GCGGCGCGCGGTGACAAACAATGGTATCGTGGTCCAGCAGCAACCACC 74089
 32 aproAlaLeaArgGluGlnAspArgLeuMetProIleAlaAsnValIleA 49
 74088 ATGTGTGCTGCTGTGACGAAGCAACATACATGCCAATCGCAACGTCATAA 74039
 49 rgIleMetArgArgValLeuProAlaHisAlaLysIleSerAspAla 65
 74038 GAATCATCGTAAACCTTACCGTCTCACGCCAAATCTCTGACGACGCC 73989
 66 LysGluThrIleGlnGluCysValSerGluTyrIleSerPheIleThrGl 82
 73988 AAGAAACGATTCAAGATGTGTCTCGAGTACATCAGCTTCGTGACCGG 73939
 82 yGluAlaasnGluArgCysGlnArgGluGlnArgLysThrIleThrAlaG 99
 73938 TGAAGCCAAACGAGCGTGTCCCAACGTCAGCAACGTAAGACCACTACTGCTG 73889
 99 luAspValLeuTrpAlaMetSerArgLeuGlyPheAspAspTyrValGlu 115
 73888 AAGATATCTTTGGCTATGACGAAGCTTGGTTCGATACACTAGCTGGAC 73839
 116 ProLeuGlyAlaTyrLeuHisArgTyrArgGluPheGluGlyAspAlaAr 132
 73838 CCCCTCACCGTGTTCATTAAACGCTACCGTACGATAGACACCGAT...CG 73792
 132 gGlyValGlyLeuValProGlyAlaAlaProSerArgGlyGlyAspHis 149
 73791 TGGTCTGCACCTT.....AGAGGT..... 73773
 149 isProHisSerMetSerProAlaAlaMetLeuLysSerArgGlyProVal 165
 73772GAGCCACCGCTGTGACACAAACC..... 73749
 166 SerGlyAlaAlaMetLeuProHisHisHisHisHisAspMetGlnMe 182
 73749 73749
 182 thisAlaAlaMetTyrGlyGlyThrAlaValProProProAlaGlyProp 199
 73748TATGAGGAATGGTATT..... 73731
 199 roHisHisGlyGlyPheLeuMetProHisProGlnGlySerSerHisTyr 215
 73730GGGTTT.....CAGGCCCATCTCATGGC 73707
 216 Leu.....ProTyrAlaTyrGluProThrTyrGlyGlyGluHi 228
 73706 CTACCTCCTCCGGTCTCTTATGTTAT.....GGTATGTT 73672
 228 saIleAlaAlaTyrTyrGlyGlyAlaAlaTyrAlaProGlyAsnGlyG 245
 73671 GGACCAATCCATGGTTATGAGAGGTGCTGCTACTACCAA...AACGGGT 73625

CDS

alignment_scores:
 Quality: 494.00 Length: 249
 Ratio: 3.146 Gaps: 9
 Percent Similarity: 63.052 Percent Identity: 46.988
 alignment_block:
 US-09-435-054-2 x AC013482/rev ..
 Align seg 1/1 to reverse of: AC013482 from: 1 to: 82875
 16 SerAlaLaGlyGlyAlaAsnGlyGlyAlaLaGlnGlnHisAlaAla 32
 74138 GCGGCGCGCGGTGACAAACAATGGTATCGTGGTCCAGCAGCAACCACC 74089
 32 aproAlaLeaArgGluGlnAspArgLeuMetProIleAlaAsnValIleA 49
 74088 ATGTGTGCTGCTGTGACGAAGCAACATACATGCCAATCGCAACGTCATAA 74039
 49 rgIleMetArgArgValLeuProAlaHisAlaLysIleSerAspAla 65
 74038 GAATCATCGTAAACCTTACCGTCTCACGCCAAATCTCTGACGACGCC 73989
 66 LysGluThrIleGlnGluCysValSerGluTyrIleSerPheIleThrGl 82
 73988 AAGAAACGATTCAAGATGTGTCTCGAGTACATCAGCTTCGTGACCGG 73939
 82 yGluAlaasnGluArgCysGlnArgGluGlnArgLysThrIleThrAlaG 99
 73938 TGAAGCCAAACGAGCGTGTCCCAACGTCAGCAACGTAAGACCACTACTGCTG 73889
 99 luAspValLeuTrpAlaMetSerArgLeuGlyPheAspAspTyrValGlu 115
 73888 AAGATATCTTTGGCTATGACGAAGCTTGGTTCGATACACTAGCTGGAC 73839
 116 ProLeuGlyAlaTyrLeuHisArgTyrArgGluPheGluGlyAspAlaAr 132
 73838 CCCCTCACCGTGTTCATTAAACGCTACCGTACGATAGACACCGAT...CG 73792
 132 gGlyValGlyLeuValProGlyAlaAlaProSerArgGlyGlyAspHis 149
 73791 TGGTCTGCACCTT.....AGAGGT..... 73773
 149 isProHisSerMetSerProAlaAlaMetLeuLysSerArgGlyProVal 165
 73772GAGCCACCGCTGTGACACAAACC..... 73749
 166 SerGlyAlaAlaMetLeuProHisHisHisHisHisAspMetGlnMe 182
 73749 73749
 182 thisAlaAlaMetTyrGlyGlyThrAlaValProProProAlaGlyProp 199
 73748TATGAGGAATGGTATT..... 73731
 199 roHisHisGlyGlyPheLeuMetProHisProGlnGlySerSerHisTyr 215
 73730GGGTTT.....CAGGCCCATCTCATGGC 73707
 216 Leu.....ProTyrAlaTyrGluProThrTyrGlyGlyGluHi 228
 73706 CTACCTCCTCCGGTCTCTTATGTTAT.....GGTATGTT 73672
 228 saIleAlaAlaTyrTyrGlyGlyAlaAlaTyrAlaProGlyAsnGlyG 245
 73671 GGACCAATCCATGGTTATGAGAGGTGCTGCTACTACCAA...AACGGGT 73625

CDS

alignment_scores:
 Quality: 494.00 Length: 249
 Ratio: 3.146 Gaps: 9
 Percent Similarity: 63.052 Percent Identity: 46.988
 alignment_block:
 US-09-435-054-2 x AC013482/rev ..
 Align seg 1/1 to reverse of: AC013482 from: 1 to: 82875
 16 SerAlaLaGlyGlyAlaAsnGlyGlyAlaLaGlnGlnHisAlaAla 32
 74138 GCGGCGCGCGGTGACAAACAATGGTATCGTGGTCCAGCAGCAACCACC 74089
 32 aproAlaLeaArgGluGlnAspArgLeuMetProIleAlaAsnValIleA 49
 74088 ATGTGTGCTGCTGTGACGAAGCAACATACATGCCAATCGCAACGTCATAA 74039
 49 rgIleMetArgArgValLeuProAlaHisAlaLysIleSerAspAla 65
 74038 GAATCATCGTAAACCTTACCGTCTCACGCCAAATCTCTGACGACGCC 73989
 66 LysGluThrIleGlnGluCysValSerGluTyrIleSerPheIleThrGl 82
 73988 AAGAAACGATTCAAGATGTGTCTCGAGTACATCAGCTTCGTGACCGG 73939
 82 yGluAlaasnGluArgCysGlnArgGluGlnArgLysThrIleThrAlaG 99
 73938 TGAAGCCAAACGAGCGTGTCCCAACGTCAGCAACGTAAGACCACTACTGCTG 73889
 99 luAspValLeuTrpAlaMetSerArgLeuGlyPheAspAspTyrValGlu 115
 73888 AAGATATCTTTGGCTATGACGAAGCTTGGTTCGATACACTAGCTGGAC 73839
 116 ProLeuGlyAlaTyrLeuHisArgTyrArgGluPheGluGlyAspAlaAr 132
 73838 CCCCTCACCGTGTTCATTAAACGCTACCGTACGATAGACACCGAT...CG 73792
 132 gGlyValGlyLeuValProGlyAlaAlaProSerArgGlyGlyAspHis 149
 73791 TGGTCTGCACCTT.....AGAGGT..... 73773
 149 isProHisSerMetSerProAlaAlaMetLeuLysSerArgGlyProVal 165
 73772GAGCCACCGCTGTGACACAAACC..... 73749
 166 SerGlyAlaAlaMetLeuProHisHisHisHisHisAspMetGlnMe 182
 73749 73749
 182 thisAlaAlaMetTyrGlyGlyThrAlaValProProProAlaGlyProp 199
 73748TATGAGGAATGGTATT..... 73731
 199 roHisHisGlyGlyPheLeuMetProHisProGlnGlySerSerHisTyr 215
 73730GGGTTT.....CAGGCCCATCTCATGGC 73707
 216 Leu.....ProTyrAlaTyrGluProThrTyrGlyGlyGluHi 228
 73706 CTACCTCCTCCGGTCTCTTATGTTAT.....GGTATGTT 73672
 228 saIleAlaAlaTyrTyrGlyGlyAlaAlaTyrAlaProGlyAsnGlyG 245
 73671 GGACCAATCCATGGTTATGAGAGGTGCTGCTACTACCAA...AACGGGT 73625

CDS

alignment_scores:
 Quality: 494.00 Length: 249
 Ratio: 3.146 Gaps: 9
 Percent Similarity: 63.052 Percent Identity: 46.988
 alignment_block:
 US-09-435-054-2 x AC013482/rev ..
 Align seg 1/1 to reverse of: AC013482 from: 1 to: 82875
 16 SerAlaLaGlyGlyAlaAsnGlyGlyAlaLaGlnGlnHisAlaAla 32
 74138 GCGGCGCGCGGTGACAAACAATGGTATCGTGGTCCAGCAGCAACCACC 74089
 32 aproAlaLeaArgGluGlnAspArgLeuMetProIleAlaAsnValIleA 49
 74088 ATGTGTGCTGCTGTGACGAAGCAACATACATGCCAATCGCAACGTCATAA 74039
 49 rgIleMetArgArgValLeuProAlaHisAlaLysIleSerAspAla 65
 74038 GAATCATCGTAAACCTTACCGTCTCACGCCAAATCTCTGACGACGCC 73989
 66 LysGluThrIleGlnGluCysValSerGluTyrIleSerPheIleThrGl 82
 73988 AAGAAACGATTCAAGATGTGTCTCGAGTACATCAGCTTCGTGACCGG 73939
 82 yGluAlaasnGluArgCysGlnArgGluGlnArgLysThrIleThrAlaG 99
 73938 TGAAGCCAAACGAGCGTGTCCCAACGTCAGCAACGTAAGACCACTACTGCTG 73889
 99 luAspValLeuTrpAlaMetSerArgLeuGlyPheAspAspTyrValGlu 115
 73888 AAGATATCTTTGGCTATGACGAAGCTTGGTTCGATACACTAGCTGGAC 73839
 116 ProLeuGlyAlaTyrLeuHisArgTyrArgGluPheGluGlyAspAlaAr 132
 73838 CCCCTCACCGTGTTCATTAAACGCTACCGTACGATAGACACCGAT...CG 73792
 132 gGlyValGlyLeuValProGlyAlaAlaProSerArgGlyGlyAspHis 149
 73791 TGGTCTGCACCTT.....AGAGGT..... 73773
 149 isProHisSerMetSerProAlaAlaMetLeuLysSerArgGlyProVal 165
 73772GAGCCACCGCTGTGACACAAACC..... 73749
 166 SerGlyAlaAlaMetLeuProHisHisHisHisHisAspMetGlnMe 182
 73749 73749
 182 thisAlaAlaMetTyrGlyGlyThrAlaValProProProAlaGlyProp 199
 73748TATGAGGAATGGTATT..... 73731
 199 roHisHisGlyGlyPheLeuMetProHisProGlnGlySerSerHisTyr 215
 73730GGGTTT.....CAGGCCCATCTCATGGC 73707
 216 Leu.....ProTyrAlaTyrGluProThrTyrGlyGlyGluHi 228
 73706 CTACCTCCTCCGGTCTCTTATGTTAT.....GGTATGTT 73672
 228 saIleAlaAlaTyrTyrGlyGlyAlaAlaTyrAlaProGlyAsnGlyG 245
 73671 GGACCAATCCATGGTTATGAGAGGTGCTGCTACTACCAA...AACGGGT 73625

10

100

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT	FEATURES
1	(bases 1 to 197976)	EU Arabidopsis sequencing, project.	Direct Submission		
		Submitted (10-NAR-2000) MIPS, at the Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail: lemcke@mips.biochem.mpg.de, mayer@mips.biochem.mpg.de	Project		
		Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK, E-mail: michael.bevan@bbsrc.ac.uk			
		Information on performance of analysis and a more detailed annotation of this entry and other sequences of chromosomes 3, 4 and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/			
		this fragment has an overlap with ATCHIRIV38 at the 5' end and an overlap with ATCHIRIV40 at the 3' end.			
		Location/Qualifiers			


```

source
1..197976
/organism="Arabidopsis thaliana"
/variety="Columbia"
/db_xref="taxon:3702"
/chromosome="4"
complement(5668..5772)
/gene="AT4g14420"
/number=1
5668..7235
/gene="AT4g14420"
complement(join(5668..5772,5865..5963,6051..6140,
6238..6352,7168..7235))
/gene="AT4g14420"
/notes="strong similarity to Nicotiana tabacum ORF able to
induce HR-like lesions, PATX:G1762945
contains EST gb:AI999161.1"
/codon_start=1
/product="elicitor like protein"
/protein_id="CAB78484.1"
/db_xref="GI:7268148"
/translation="MELASELGRALFVSFLLSAWQEFNFGEDGGRSAKSLPKFNA
FNVHTTHTQQLPVDPMKILVAAAIKLGIGLLFVGSLSGLAYLILLHQAVATPIL
YDFNYDVRKEQFLSKFTQSLALLGLLFFTGKMSRKHGRQLPKKAPKAKAN"
complement(join(5668..5772,5865..5963,6051..6140,
6238..6352,7168..7235))
/gene="AT4g14420"
complement(5773..5864)
/number=1
complement(5865..5963)
/gene="AT4g14420"
/number=2
complement(5964..6050)
/number=2
complement(6051..6140)
/gene="AT4g14420"
/number=3
complement(6141..6237)
/number=3
complement(6238..6352)
/gene="AT4g14420"
/number=4
complement(6353..7167)
/number=4
complement(7168..7235)
/gene="AT4g14420"
/number=5
complement(8407..9129)
/gene="AT4g14430"
/number=1
complement(8407..9129)
/gene="AT4g14430"
complement(8407..9129)
/gene="AT4g14430"
/notes="similarity to carnitine racemase, Escherichia coli
contains EST gb:H35845, AI99737.1, T76583"
/codon_start=1
/product="carnitine racemase like protein"
/protein_id="CAB78485.1"
/db_xref="GI:7268149"
/translation="MCTLEKRGDLFLLTLTGDEHFRHFDPTTASVLSLLEQAKSQSTR
GSLITITANGKFEFNGDLAWAQTAGSKTGAANRLHOMVESKPVVAALLDLPMTTIA
ALNGHAAAGLILALSHDIVFMKRDGVLYMSEVDIGLSPDIFYSALVRKLTGTSAAR
RELLSGKILGEEAVGLGIVDSAAAYDSEGVVAVSVRLGEKLAARKWSGEVYASIRK
SLYPELGLGLETRVFATPKL"
complement(10443..11159)
/gene="AT4g14440"
/notes="similarity to carnitine racemase, Escherichia coli
contains EST gb:AI994308.1, AA394352, H36791, AA395864,
T76583, H77091, T46265, AA597888, Z17630"
/codon_start=1
/product="carnitine racemase like protein"

```

```

/protein_id="CAB78486.1"
/db_xref="GI:7268150"
/translation="MCTLEKRGDLFLLTLTGDEHFRHFDPTTASVLSLLEQAKSQSTR
GSLITITANGKFEFNGDLAWAQTAGSKTGAANRLHOMVESKPVVAALLDLPMTTIA
ALNGHAAAGLILALSHDIVFMKRDGVLYMSEVDIGLSPDIFYSALVRKLTGTSAAR
RELLSGKILGEEAVGLGIVDSAAADSAEGVVEATVSLGESLAARKWNGEVYATIRKSL
YPELCRMVDLTANNATHNL"
complement(10443..11159)
/gene="AT4g14440"
/number=1
complement(10443..11159)
/gene="AT4g14440"
complement(10443..11159)
/notes="similarity to adenylate cyclase (EC 4.6.1.1) II,
rat
contains EST gb:AF067468, AI994518.1, T44124"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAB78487.1"
/db_xref="GI:7268151"
/translation="MASTNORTSSSAASNERQLSQRRAPSLMIKPTSFNSNVAIP
LISPLAPLSTSSFDQSHVPPQNKTEIPVEEVKTPVFKWQHPASPCFYPTTENR
ASRTSLFDGLGLEGRLRASSYAHDRDNDEALENLQDRVSLFKRVCFHLYN"
13366..14295
/gene="AT4g14450"
/number=1
14296..14473
/gene="AT4g14450"
/number=2
17117..17336
17117..22563
/notes="LTR retrotransposon"
17575..22327
/gene="AT4g14460"
join(17575..19093,19145..19476,19552..21256,21338..21472,
21549..22327)
/gene="AT4g14460"
/notes="similarity to similarity to retrovirus-related
polyprotein (retrotransposon Tat-3), Arabidopsis thaliana"
/codon_start=1
/product="retrovirus-related like polyprotein"
/protein_id="CAB78488.1"
/db_xref="GI:7268152"
/translation="MGSVPPANPVTRPPINNGDQIRYPVDQVENVYVYLSADHAGL
ILVSRLLTASDFHSWRKSLMALNVRNKLGFINGITTKPEDHRDFGAWRCNDIVS
TWLMSVDKKIQGSLLYIATVQGINNLLSRFKQDDAPRIFDIQKLSKIEQSDMS
TYITALLTWEHRNVELPVCCTGCRNCCDAVKAWEHQORSVTKLSKELNEGFDOT
RRHILKPIPTIKEAFNNVTQDERQRNKLTRVDSVAFONTSMINEDENAYVAYN
TVRNPKPTCTHCKVGHTIOCKYKVGYPGKMGNTGYTKPNQLHVGPRMPMP
QPRMOPPAQVNTSMQKANYVAQYAECAIPSEGYSOAPMNNPYGTPMPTHGGN
NLSQDFTPOQIEQMSIQFQAQVQVPEAASNNPSPATVSEHGFALTSTGTTIP
FPSTSLKYENNDLQPNHTLSALQKFLPSDAWIIDSGASSHVCSDLAMFRELKSVSGT
VHTQKLLHNLVHLVDFKFLMSVSLVKTISCSAHFYVDCCLIQELSOGLMIGRGR
LYHNLYLETENTSPSTPAACLTGSLVNDHLHORLGHPSVSVLVOKLKLRLAYS
HNILASNPDLVHLDIQGPFSEIETEGFRYELTVVDDCTFTWYMLNKKDSSVFP
EFIKLVSTQFNAKIKATISDNAPELGTEIVEKGMHLHFSCTATPOONSVVERKHQH
ILNVAFLFQSNIPMQVSCVTITAVFLINRLPSLNNKSPYELILNKOPYSLLK
NFGCLFVSTNAHERKFTPRARCVFLGSPGYKGYKVLDSHSTVSRNVVFKKH
VFPKTSLLNKAQVDMFENSLPLPAPLHFVETMPLIDEDSLIPTTDSRTADNHASS
SSSALPSTIPSSNETQIDISNAVPIRTRKTRAPSLSEYHCSILVPSISTPPTD
SSIPHLPELFTASSPKKTPYPISTVSVYDKVTPLCOSYIFAYNETETPKFSQAM
KSEKIRVAVELQAMELNTWVESLPPDKNVVCGKWVFIKYNPDGTVERTYKARLV
AOGTQOQGIQDPLDTFSVAKLTSKMMGLIAAITGMTQMDVSDVSLAFLLDDDEEIF
MSLFGYTPPAGTILPPNPVCRLLKSIYGLKQASQWYKRFVAAALVYIDIMIASNND
AEVENIKALLRSEFKIKDLGPAPRFLGLLCKPSSIPMDPTLHLVRDNGTPLPNPTAY

```

Quality: 392.50 Length: 191
Ratio: 2.784 Gaps: 5
Percent Similarity: 73.822 Percent Identity: 42.932

alignment_block:

US-09-435-054-2 x ZMNFYB ..

Align seg 1/1 to: ZMNFYB from: 1 to: 870

```

6 PheLeuProAlaAlaGlyAlaGluAsnGlySerAlaAlaGly.GlyAlaA 22
||| .....:|||||:.....:|||||:.....:|||||:.....:
4 TTCCGGATAAGCGCAGGAGGAGCTCATCGCGAAGCTCCGGCGAGCCCTG 53
.....:|||||:.....:|||||:.....:|||||:.....:|||||:.....:
22 snAsnGlyGlyAlaAlaGlnGlnHisAlaAlaPro.....: 33
.....:|||||:.....:|||||:.....:|||||:.....:|||||:.....:
54 GCGGGCGCGCGGAGCCAGCAGAGCGGAGCGCCCGAGGGGCGGAGGC 103
34 .....AlaIleArgGluGlnAspArgLeuMetProIleAlaAsnValI 48
||| .....:|||||:.....:|||||:.....:|||||:.....:
104 GTGGCAGCGCTCAGGAGCAGCAGAGGTTCCTGCCCAACATCATG 153
48 eArgIleMetArgAlaValLeuProAlaHisAlaLysIleSerAspAspA 65
||||| .....:|||||:.....:|||||:.....:|||||:.....:|||||:.....:
154 TCGCATCATGAGAGGCGCATCCCGCTTAACCGGAGATCGCCCAAGGAG 203
65 lAlaGluThrIleGlnGluCysValSerGluTyrIleSerPheIleThr 81
||||| .....:|||||:.....:|||||:.....:|||||:.....:|||||:.....:
204 CTAAGGAGACCGTGCAGGAGTGCCTCGAGTTCCTCTCATCACT 253
82 GlyGluAlaAsnGluArgCysGlnArgGluGlnArgLysThrAl 98
.....:|||||:.....:|||||:.....:|||||:.....:|||||:.....:
254 AGCGAAGCAGTGCACAGTGCAGAGGAGGAGGAGGAGGAGGAGGAGG 303
98 aGluAspValLeuTrpAlaMetSerArgLeuGlyPheAspTyrValG 115
.....:|||||:.....:|||||:.....:|||||:.....:|||||:.....:
304 CGACGATCTGCTGTGGCCATCGCCACGCTGGGGTTGAAGACTACATG 353
115 luProLeuGlyAlaTyrLeuHisArgTyrArgGluPheGluGlyAspAla 131
||||| .....:|||||:.....:|||||:.....:|||||:.....:|||||:.....:
354 AACCCCTCAAGGTACCTACAGAGTACAGAGAGATGAGGAGGTGATAGC 403
132 ArgGlyValGlyLeuValProGlyAlaAlaProSerArgGlyGlyAsp.. 147
AA .....:|||||:.....:|||||:.....:|||||:.....:
404 AAG.....TTAAGTCTAAATCTAGCGATGGCTCGAT 435
148 .....HisHisProHisSerMetSerProAlaAlaM 158
||||| .....:|||||:.....:|||||:.....:|||||:.....:
436 TAAAGAGATGCTCTGTGTCATGTGGGAGCAAGTACCTCAGCTGCAGAAG 485
158 eLeuLysSerArgGlyProValSer...GlyAlaAlaMetLeuProHis 173
:: .....:|||||:.....:|||||:.....:|||||:.....:
486 GGATGGCCACAGGAGGAGCATACACCAAGGAATGGGTATATATGCAACCT 535
174 HisHisHisHisAspMet 180
.....:|||||:.....:|||||:.....:|||||:.....:
536 CAGTACCATACGGGATATC 556

```

seq_name: gb_p12:ATHAP3A

```

seq_documentation_block: 832 bp mRNA PLN 12-SEP-1997
LOCUS ATHAP3A
DEFINITION Arabidopsis thaliana mRNA for Hap3a transcription factor.
ACCESSION Y13723
VERSION Y13723.1 GI:2398526
KEYWORDS hap3a gene; transcription factor.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta;
Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 832)
AUTHORS Edwards,D.
TITLE Direct Submission

```

JOURNAL

Submitted (12-JUN-1997) Edwards D., Plant Sciences, University of

Cambridge, Downing Street, Cambridge CB2 3EA UK

2 (bases 1 to 832)

Edwards,D., Smith,A.G. and Murray,J.A.

Isolation and Characterisation of CCAAT box binding proteins from

higher plants

Unpublished

FEATURES

Location/Qualifiers

1..832

/organism="Arabidopsis thaliana"

/db_xref="taxon:3702"

103..528

/gene="hap3a"

103..528

/gene="hap3a"

/codon_start=1

/product="transcription factor"

/protein_id="CAA74051.1"

/db_xref="GI:2398527"

/translation="MADTPSPAGSGSGSVREQDRLPIANISIMKKALPPNGK

IGKDAKTVQECVSEFISFITSSEADKCKEKRKTVNGDILLWAMATLGFEDYLEPLK

IYLARYLEEDGDKSGSGKSGDGSNRDAGSGVSGEEMPSM"

BASE COUNT 252 a 132 c 204 g 244 t

ORIGIN

alignment_scores: Quality: 376.50 Length: 163

Ratio: 3.303 Gaps: 3

Percent Similarity: 69.939 Percent Identity: 46.012

alignment_block:

US-09-435-054-2 x ATHAP3A ..

Align seg 1/1 to: ATHAP3A from: 1 to: 832

```

8 ProAlaAlaGlyAlaGluAsnGlySerAlaAlaGlyGlyAlaAsnAsnG 24
||||| .....:|||||:.....:|||||:.....:|||||:.....:
115 CCTTCGAGCCAGCTGGAGATGGCGGAGAAAGCGCGT..... 153
24 yGlyAlaAlaGlnGlnHisAlaAlaProAlaIleArgGluGluAspArgL 41
||||| .....:|||||:.....:|||||:.....:|||||:.....:
154 .....TCCGTTAGGAGCAGGATCGAT 175
41 euMetProIleAlaAsnValIleArgIleMetArgArgValLeuProAla 57
||||| .....:|||||:.....:|||||:.....:|||||:.....:
176 ACCTTCCTATAGCTAATATCAGCAGGATCATGAAGAAAGCGTTGCCCTCT 225
58 HisAlaLysIleSerAspAlaLysGluThrIleGlnGluCysValSe 74
||||| .....:|||||:.....:|||||:.....:|||||:.....:
226 AATGGTAAGATTGAAAGATGCTAAGGATACAGTTCAGGAATCGCGTCTC 275
74 rGluTyrIleSerPheIleThrGlyGluAlaAsnGluArgCysGlnArg 91
||||| .....:|||||:.....:|||||:.....:|||||:.....:
276 TGAGTTTCATCAGCTTCATCAGTACGAGCGCAGTGAAGTGTCAGGAGG 325
91 luGlnArgLysThrIleThrAlaGluAspValLeuTrpAlaMetSerArg 107
||||| .....:|||||:.....:|||||:.....:|||||:.....:
326 AGAAAGAGAAACCTGTAATGTCATGATTTGTTGGCAATGGCAACA 375
108 LeuGlyPheAspAspTyrValGluProLeuGlyAlaTyrLeuHisArgTy 124
||||| .....:|||||:.....:|||||:.....:|||||:.....:
376 TTAGGATTGAGGATTACCTGGAACCTCTAAAGATATACCTAGCGAGGTA 425
124 rArgGluPheGluGlyAspAlaArgGlyValGlyLeuValProGlyAla 141
||||| .....:|||||:.....:|||||:.....:|||||:.....:
426 CAGGAGTTGAGGCTGATTAATAGCGATCAGGA..... 459
141 laProSerArgGlyGlyAspHisHisProHisSerMetSerProAlaAla 157
||||| .....:|||||:.....:|||||:.....:|||||:.....:
460 .....AAGAGTCGAGAT.....GGATCA 477

```

158 MetLeuylsSerArgGlyProValSerGlyAlaMet 170
 478 AATAGACATGCTGGTGGGTGTTCTCTGGTGAAGAAATG 516

seq_name: gb_pl1:AC005309

seq_documentation_block:
 LOCUS AC005309 107377 bp DNA PLN 05-APR-2000
 DEFINITION Arabidopsis thaliana chromosome II section 254 of 255 of the
 complete sequence. Sequence from clones T30B22, F17A22, T9J23.
 AC005309 AE02093
 VERSION AC005309.2 GI:6598453
 KEYWORDS HTG.
 SOURCE thale cress.
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 Magnoliophyta; eudicotyledons: core eudicots; Rosidae; eurosids II;
 Brassicales; Brassicaceae; Arabidopsis.

REFERENCE
 AUTHORS 1. (bases 1 to 107377)
 Lin, X., Kaul, S., Rounsley, S.D., Shea, T.P., Benito, M.-I., Town, C.D.,
 Fujii, C.Y., Mason, T.M., Bowman, C.L., Barnstead, M.E.,
 Feldblyum, T.V., Buell, C.R., Ketchum, K.A., Lee, J.J., Ronning, C.M.,
 Koo, H., Moffat, K.S., Cronin, L.A., Shen, M., VanAken, S.E., Umayam, L.,
 Tallon, L.J., Gill, J.E., Adams, M.D., Cartera, A.J., Creasy, T.H.,
 Goodman, H.M., Somerville, C.R., Copenhaver, G.P., Preuss, D.,
 Nierman, W.C., White, O., Eisen, J.A., Salzberg, S.L., Fraser, C.M. and
 Venter, J.C.

TITLE Sequence and analysis of chromosome 2 of the plant Arabidopsis

JOURNAL thaliana
 MEDLINE Nature 402 (6763), 761-768 (1999)
 PUBMED 20093487

REFERENCE 10617197

AUTHORS 2. (bases 1 to 107377)

TITLE Lin, X.

JOURNAL Direct Submission

COMMENT Submitted (09-MAR-2000) The Institute for Genomic Research, 9712
 Medical Center Dr., Rockville, MD 20850, USA
 On Dec 17, 1999 this sequence version replaced gi:3738275.
 The sequence and annotation of chromosome 2 were merged from those
 of the individual clones on this chromosome after removing
 overlaps. For detailed information, please see the TIGR web site
 (<http://www.tigr.org/tdb/at/at.html>).

Genes were identified by a combination of three methods: Gene
 prediction programs including GRAIL
 (<ftp://arthur.epm.ornl.gov/pub/xgrail>), GeneFinder (Phil Green,
 University of Washington), Genscan (Chris Burge,
<http://gnomic.stanford.edu/GENSCANW.html>), and NetPlantGene
 (<http://www.cbs.dtu.dk/services/NetGene2/>), searches of the
 complete sequence against a peptide database and plant EST
 databases at TIGR, and manual curations based on those analyses.
 Annotated genes are named to indicate the level of evidence for
 their annotation. Genes with similarity to other proteins are named
 after the database hits. Genes without significant peptide
 similarity but with EST similarity are named as 'unknown' proteins.
 Genes without protein or EST similarity, that are predicted by two
 or more gene prediction programs over most of their length are
 annotated as 'hypothetical' proteins. Genes encoding tRNAs are
 predicted by tRNAscan-SE (Sean Eddy,
<http://genome.wustl.edu/eddy/tRNAscan-SE/>). Simple repeats were
 identified by RepeatMasker (Arian Smit,
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>). Genes are
 numbered from the top to bottom of the chromosome.

We thank the CSHL/WashU/ABI consortium for sequencing BAC clones
 F6P23, F5J6, T1A5, and T1316, the ESSA group for sequencing clone
 F13B4, and Scott Jackson, Jiming Jiang, Klaus Meyer, Eric Richards
 and Satoshi Tabata for helpful assistance. In addition, we would
 like to thank the TIGR Bioinformatics Department, especially Lixin
 Zhou, Hanif Khalak, Michael E. Heaney, Lily Fu, Feng Liang, Jeremy
 Peterson, Michael Holmes, and Delwood Richardson for software and
 database support.

This work was supported by the National Science Foundation,
 Department of Energy and the US Department of Agriculture.

Address all correspondence to: atetigr.org.

FEATURES
 source

Location/Qualifiers

1..107377

/organism="Arabidopsis thaliana"

/cultivar="Columbia"

/db_xref="taxon:3702"

/chromosome="II"

complement(<1..821)

/note="Sequence from clone T30B22"

join(<291..348,504..1024,1120..1549,2112..2233,

2332..>2508)

/gene="At2g47620"

<291..>2508

/gene="At2g47620"

/note="T30B22.7"

join(291..348,504..1024,1120..1549,2112..2233,2332..2508)

/gene="At2g47620"

/codon_start=1

/product="putative SWI/SNF family transcription activator"

/protein_id="AAC63618.2"

/db_xref="GI:6598454"

/translation="MEATDPSAEIETYTPAQSSWFLWDDIHEIRERFEAFPTESI

TRTPKVDKLYSDVNAKTEOCTPAGIRVATPNSLRPTAPPLVEERVETGKVPPLTS

YSDFSLKKPDHVLCAHGCERCDSPFYOHNGKIVNICCKFKNGNGYNNATDFK

LIGNSAAVNTREIELLLLSVLKHGDWELISQSVSTKSLDCISKLIELPGEFLM

GSASGRNPSLTEDENTEQVQDQHEETSTREEDKRVNDEPPAKRKRVALISE

SDTLIRASVATLGAQAQALIAQEREMEQLAATVIEQQLKQSLKFLDDILE

SIMDEEEKVIEGVKETIIQERSVVLQCAFRTGKITKRDHTYK"

822..107176

/note="Sequence from clone F17A22"

complement(1028..1102)

/rpt_family="(CAAAA)n"

1216..1715

/gene="At2g47620"

/note="molecular marker m179a"

join(<3826..4057,4175..4305,4398..4535,4614..4712,

4788..>5168)

/gene="At2g47630"

<3826..>5168

/gene="At2g47630"

/note="F17A22.2"

join(3826..4057,4175..4305,4398..4535,4614..4712,

4788..5168)

/gene="At2g47630"

/codon_start=1

/product="putative phospholipase"

/protein_id="AAC63619.1"

/db_xref="GI:3738277"

/translation="MASLVVEKIYSEVDTCKLNKLHVLICGIRLASAGAVFGM

DYEGHGRMSRCYIKKFNVDYITSCAQEYDKGRFLYGESMGAVTLIL

HKKDPFNWAGNLIYAPCKISERKPHPIVNLITRVETIIPKVIPTKVDVDAAFK

DLVKREVRNKLITYQKPRKLTALEMLRTSMNLEDTLHEITPFFVLHGEADTVDP

EVSKALYKASTDKTKLKYPMWHALTSQPCNDVDFADLIINWLDLTADPASIT

VTPLRVGNTTSVORVTTVGVSNGHREPKRPFFNLLGGLNGLVPRSTV"

join(5465..5512,5607..5654,5762..5786,6120..6194,

6385..6489,6572..6844)

/gene="At2g47640"

5465..6844

/gene="At2g47640"

/note="F17A22.3"

join(5650..5654,5762..5786,6120..6194,6385..6489,

6572..6688)

/gene="At2g47640"

/codon_start=1

/product="putative small nuclear ribonucleoprotein D2"

/protein_id="AAC63620.1"

/db_xref="GI:3738278"

/translation="MSKPMBEDTNGKTEEEFNTPGLPSVLMSVKNNTQVLINCRNR

KLLGRVRAFRHCHNMVLENVREMTVEPKTGKGGKALPVNRDRFISKMFGRGDSVII

18 AlaGlyGlyAlaAsnAsnGlyGlyAlaAla.GlnGlnHisAlaAlaProA 34

```
repeat_region complement(34637..34727)
mRNA /rpt_family="(TA)n"
gene /complement(38932..40224)
     /gene="At2g13550"
     /complement(38932..40224)
     /gene="At2g13550"
     /note="T10F5.9"
CDS complement(39231..39776)
     /gene="At2g13550"
     /note="unknown protein"
     /codon_start=1
     /protein_id="AAD22678.1"
     /db_xref="GI:4558660"
     /translation="MTSQIRQTPLARHMLTQFPNTRGRQKSPNSHHDPDESPS
POQRHITREPTIRGTHQKLTSSSTANTPTGDKRQKHIVTNTDIASDKTRS
RSGNQSPQPHIQKLSKLTGAANNIETNLNRRYRREIGRPTQHQHPYNNHITKT
RRKSKDEIKORTRPSVLK"
mRNA join(<40516..40686,40766..40862,40943..41061,41171..41338,
41499..41627,41718..41778,42210..42313,42390..42473,
alignment_scores:
  Quality: 354.50 Length: 209
  Ratio: 2.686 Gaps: 6
Percent Similarity: 63.158 Percent Identity: 37.799
alignment_block:
US-09-435-054-2 x AC007063/rev ..
Align seg 1/1 to reverse of: AC007063 from: 1 to: 82403
8 ProAlaAlaGlyAlaGluAsn...GlySerAlaAlaGlyGlyAlaAsnAs 23
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
46880 CTGGAGTAGCTGAACAATAATCCAGGAAGCCCTTCTTCAAGACCAACAA 46831
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
23 nGlyGlyAlaAlaGlnGlnHisAlaProAlaIleArgGluGlnAspA 40
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
46830 CAACACACACACAC.....AAAGACAGGACC 46802
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
40 rGluMetProIleAlaAsnValIleArgIleMetArgValLeuPro 56
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
46801 GGTTCCTCCCATGCGAATGCGGAAGGATCATGAAAGATGTTCTTCCC 46752
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
57 AlaHisAlaLysIleSerAspAlaLysGluThrIleGlnGluCysVa 73
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
46751 GGTAACGGTAAGATCTCAAAAGACGCTAAAGAACCGTTCAGAAATGTGT 46702
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
73 lSerGluTyrIleSerPheIleThrGlyGluAlaAsnGluArgCysGlnA 90
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
46701 CTGGAGTAGCTAGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 46652
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
90 rGluGlnArgLysThrIleThrAlaGluAspValLeuTrpAlaMetSer 106
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
46651 GAGAAAGAGAGAGAGACCATCAATGGAGATGATATCTTTGGGCTATCACA 46602
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
107 ArgLeuGlyPheAspAspTyrValGluProLeuGlyAlaTyrLeuHisAr 123
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
46601 ACTCTCGGTTTCCGAGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 46552
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
123 gTyrArgGluPheGluGlyAspAlaArgGlyValGlyLeuValProGlyA 140
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
46551 ATATAGACACCGAAGGAGAGAAAGTTAAC..... 46521
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
140 lAlaProSer.....ArgGly 145
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
46520 ..AGCCAAACAAACAAACAAACAAACAAACAAACAGCAGATTCAACAA 46473
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
146 GlyAspHisHisProHisSerMetSerProAlaAlaMetLeuLysSerAr 162
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
46472 CAGAATCATCAATAATTATCAGTTTCAAGAACCAAGAC.....CAAAACAA 46429
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
162 gGlyProValSerGlyAlaAlaMetLeuProHisHisHisHisHisA 179
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
46428 TAACACATGTCTACTAGTTACATCTCTCATCATCAT..... 46389
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
```

```
179 spMetGlnMetHisAlaAlaMetTyrGlyGlyThrAlaValProProPro 195
46388 .....CCTTCTCCA 46380
196 AlaGlyProProHisHisGlyGlyPhe 204
|||||
46379 TTCTACCATGGATCATCAACCTTTT 46353
seq_name: gb_ov:AF041204
seq_documentation_block:
LOCUS AF041204 787 bp mRNA VRT 02-DEC-1998
DEFINITION Xenopus laevis nuclear Y/CCAAT-box binding factor B subunit NF-YB
mRNA, complete cds.
ACCESSION AF041204
VERSION AF041204.1 GI:3170224
KEYWORDS African clawed frog.
SOURCE Xenopus laevis
ORGANISM Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
Xenopodinae; Xenopus.
REFERENCE 1 (bases 1 to 787)
AUTHORS Li,Q., Herrler,M., Landsberger,N., Kaludov,N., Ogrzykzo,V.V.,
Nakatani,Y. and Wolffe,A.P.
TITLE Xenopus NF-Y pre-sets chromatin to potentiate p300 and
acetylation-responsive transcription from the Xenopus hsp70
promoter in vivo
JOURNAL EMBO J. 17 (21), 6300-6315 (1998)
MEDLINE 99016050
REFERENCE 2 (bases 1 to 787)
AUTHORS Herrler,M. and Wolffe,A.P.
TITLE Direct Submission
JOURNAL Submitted (07-JAN-1998) Laboratory of Molecular Embryology,
National Institutes of Health/NICHD, Building 18T, Room 106,
Bethesda, MD 20892, USA
FEATURES
Location/Qualifiers
1..787
/organism="Xenopus laevis"
/db_xref="taxon:8355"
/cell_type="oocyte"
46..666
/function="transcription factor"
/note="similar to CBP-A and yeast HAP-3"
/codon_start=1
/product="nuclear Y/CCAAT-box binding factor B subunit
NF-YB"
/protein_id="AAC82336.1"
/db_xref="GI:3170225"
/translation="MDGDSATTTDTLSQIGSDYISGSHYVIQPHDGDTSINDPDDT
NGSKDSFREQDIYLPINAVRIMKNVPTGKIADKAKECVQECVSEFISFTSEASE
RCHOEKRTINGEDILFAMSLRGFDSYVEPLKLYLQKPRAMKKGIGCTVTTGDGL
GEDLAEPSTSOIPAGLITDGOQONVMVYTSYQQISGVQIQFS"
BASE COUNT 263 a 149 c 187 g 188 t
ORIGIN
```

```
alignment_scores:
  Quality: 340.50 Length: 145
  Ratio: 2.987 Gaps: 3
Percent Similarity: 78.621 Percent Identity: 49.655
alignment_block:
US-09-435-054-2 x AF041204 ..
Align seg 1/1 to: AF041204 from: 1 to: 787
34 AlaIleArgGluGlnAspArgLeuMetProIleAlaAsnValIleArgI 50
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
193 AGTTTCAGAGAGCAAGATATTATCTTCCCATCGCTAATCTGCGAAGGAT 242
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
50 eMetArgArgValLeuProAlaHisAlaLysIleSerAspAlaLysG 67
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
```

```

|||||.....:|||||.....:|||||.....:|||||.....:|||||.....:
243 AATGAAATCTGCTACACAGCAGGAAATTTGCAAAAGACGCAAAAG 292

67 luThrIleGlnGluCysValSerGluThrIleSerPheIleThrGlyGlu 83
|| .....:|||||.....:|||||.....:|||||.....:|||||.....:
293 AGTGTGTACAGGAATGCGTAAGCGAATTTATAGCTTCAATCAATCGAA 342

84 AlaAsnGluArgCysGlnArgGluGlnArgGluThrIleThrAlaGluAs 100
|||||.....:|||||.....:|||||.....:|||||.....:|||||.....:
343 GCAAGTGGAGGATGCCACAGGAGGAGCAAGAACCAATCAATGAGAGA 392

100 pValLeuTrpAlaMetSerArgLeuGlyPheAspTyrValGluProL 117
|| .....:|||||.....:|||||.....:|||||.....:|||||.....:
393 TATTCATTATGCTATGCGAGACTAGGCTTTGATAGTATGATAGCCAT 442

117 euGlyAlaTyrLeuHisArgTyrArgGluPheGluGlyAspAlaArgGly 133
|| .....:|||||.....:|||||.....:|||||.....:|||||.....:
443 TAAAGTTGTATCTTCAGAAATTCAGAGAGCAATGAAAGGGGAAAGGC 492

134 ValGlyLeuValProGlyAlaAlaProSerArgGlyGlyAsp.....HI 148
:|||||.....:|||||.....:|||||.....:|||||.....:|||||.....:
493 ATGTGCTGTACAGTTACACAGAGATGGTCTGGCGCAAGACCTTCGAGA 542

148 sHisProHis...SerMetSerProAlaAlaMetLeuLysSerArgGlyP 164
:|||||.....:|||||.....:|||||.....:|||||.....:|||||.....:
543 AGAACCTTTCACTAGCAATACCAAGCAGGTTTAAATACCAAGATGGAC 592

164 roValSerGlyAlaAlaMetLeuProHisHis 174
|||||.....:|||||.....:|||||.....:|||||.....:|||||.....:
593 AGCAACAGAGATGTTATGGTTTATACCATCAT 625

```

seq_name: gb_ov:PMNFBYB

```

seq_documentation_block: 699 bp mRNA VRT 09-FEB-1999
LOCUS PMNFBYB
DEFINITION P.marinus mRNA for CAAT-box DNA binding protein subunit B (NF-YB).
ACCESSION X59712
VERSION X59712.1 GI:64217
KEYWORDS CAAT-box DNA binding protein.
SOURCE sea lamprey.
ORGANISM Petromyzon marinus

```

```

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Hyperoartia;
AUTHORS Petromyzontiformes; Petromyzontidae; Petromyzon.
TITLE 1 (bases 1 to 699)
JOURNAL Submitted (21-JAN-1992) C. Benoist, L.G.M.E., Dept of Immunology,
11, Rue Humann, Strassbourg 67000, FRANCE
AUTHORS Li.X.Y., Mantovani,R., Hooft van Huijsdijnen,R., Andre,I.,
Benoist,C. and Mathis,D.
TITLE Evolutionary variation of the CCAAT-binding transcription factor
NF-Y
JOURNAL Nucleic Acids Res. 20 (5), 1087-1091 (1992)
MEDLINE 92195809
REMARK Erratum: [[published erratum appears in Nucleic Acids Res 1992 Apr
11;20(7):184]]

```

```

FEATURES
source 1..699
/organism="Petromyzon marinus"
/db_xref="taxon:7757"
40..669
/codon_start=1
/product="CAAT-box DNA binding protein subunit B (NF-YB)"
/protein_id="CA44232.1"
/db_xref="GI:64218"
/db_xref="SWISS-PROT:P25210"
/translaton="MDADGSTTDASOLGITGDYIGGGHVLOSSDGDAGSLASGDHD
ESGSKDPYEQDIYLPANVARIMKTSIPSSGKIADKAECVQECVSEIFISFSEA
SERCHQEKRTINGEDILFAMSTLGFDSVPELKYQVLRSMKGEKGINAVVITTT
DAIPELTESFSGPLATSLITADGQQQNVMTYTAIQIGVQPIQFT"
BASE COUNT 186 a 179 c 205 g 129 t
ORIGIN

```

```

alignment_scores:
Quality: 325.50 Length: 124
Ratio: 3.356 Gaps: 1
Percent Similarity: 78.226 Percent Identity: 53.226

alignment_block:
US-09-435-054-2 x PMNFBYB ..
Align seg 1/1 to: PMNFBYB from: 1 to: 699

11 GlyAlaGluAsnGlySerAlaAlaGlyGlyAlaAsnAsnGlyGlyAlaAl 27
|| .....:|||||.....:|||||.....:|||||.....:|||||.....:
133 GGTGATGCTGAGGTAGCTGCGAGTGGTGTACGATGAGAGCTGTGG 182

27 agLInGlnHisAlaAlaProAlaIleArgGluGlnAspArgLeuMetProI 44
:|||||.....:|||||.....:|||||.....:|||||.....:|||||.....:
183 ATCGAAA.....GACCCGCTATCGCGAGCAGCACATCTACCTGCCGA 223

44 leAlaAsnValIleArgIleMetArgArgValLeuProAlaHisAlaLys 60
|||||.....:|||||.....:|||||.....:|||||.....:|||||.....:
224 TAGCTAAGCTGCCCGGATCATGAAGACCTCATCCCATCTCTGGGAG 273

61 IleSerAspAlaLysGluThrIleGlnGluCysValSerGluTyrI 77
|||||.....:|||||.....:|||||.....:|||||.....:|||||.....:
274 ATTGCAAGGACGCGCAAGGAGTGTGTGCGAGGAGTGCCTGAGCGAATTCAT 323

77 eSerPheIleThrGlyGluAlaAsnGluArgCysGlnArgGluGlnArgL 94
|||||.....:|||||.....:|||||.....:|||||.....:|||||.....:
324 CAGCTTCATCATCGTCCGGAAGCGAGCGACGCTGCCACGAGGAAACGCA 373

94 ysThrIleThrAlaGluAspValLeuTrpAlaMetSerArgLeuGlyPhe 110
|||||.....:|||||.....:|||||.....:|||||.....:|||||.....:
374 AGACCATCAATGGGAGGAGCATCTCTTCGCGATGTCCACCTCTGGCTTC 423

111 AspAspTyrValGluProLeuGlyAlaTyrLeuHisArgTyrArgGluPh 127
|||||.....:|||||.....:|||||.....:|||||.....:|||||.....:
424 GACAGTACGTGCGAGCGCTCAAGCAGCATCTTCAAAAGTACCGTGAGTC 473

127 eGluGlyAspAlaArgGlyVal 134
|||||.....:|||||.....:|||||.....:|||||.....:|||||.....:
474 AATGAAGGCTGAGAAAGGAATC 495

```

seq_name: gb_ro2:RATCBFALL

```

seq_documentation_block: 538 bp mRNA ROD 27-APR-1993
LOCUS RATCBFALL
DEFINITION Rat CCAAT binding transcription factor-B subunit (CBF-B1) mRNA,
complete cds.
ACCESSION M60617 J05701
VERSION M60617.1 GI:203354
KEYWORDS CCAAT binding transcription factor-B subunit.
SOURCE Rat (Sprague-Dawley) liver, cdna to mRNA.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 538)
AUTHORS Vuorio,T., Maiti,S.N. and de Crombrughe,B.
TITLE Purification and molecular cloning of the A chain of a rat
heteromeric CCAAT-binding protein: Sequence identity with the yeast
HAP3 transcription factor
JOURNAL J. Biol. Chem. 265, 22480-22486 (1990)
MEDLINE 91093096
FEATURES
Location/Qualifiers
1..538
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/tissue_type="liver"
/tissue_lib="lambda-ZAP"
1..447
gene

```

```
CDS
1..447
/gene="CBF-All"
/gene="CBF-All"
/codon_start=1
/product="CCAAT binding transcription factor-B subunit"
/protein_id="AAA40888.1"
/db_xref="GI:203355"
/translation="MNDHEDTNGSKSFREQDIYLPITANVARIMKNAIPOTGKIADAK
KEQVECVSEFISITSEASERCHOEKRTINGEDILFAMSTLGFDSYVEPLKLYLOK
FREAMKGEKGGGAVSATDGLSELTETAEFSKLGTAAAGEFSPWAVC"
BASE COUNT 158 a 106 c 152 g 122 t
ORIGIN
```

```
alignment_scores:
  Quality: 324.50      Length: 113
  Ratio: 3.527        Gaps: 1
  Percent Similarity: 81.416 Percent Identity: 56.637

alignment_block:
US-09-435-054-2 x RATCBFAll ..
Align seg 1/1 to: RATCBFAll from: 1 to: 538

34 AlaIleArgGluGlnAspArgLeuMetProIleAlaAsnValIleArgI1 50
   ::::|||||
37 AGTTTCAGAGAACAGATATTATCTCCCATTCGCAATCTGGCTAGGAT 86
   :|||
50 eMetArgValLeuProAlaHisAlaLysIleSerAspAlaLysG 67
   |||
87 AATGAAATGCCATACCTCAACAGGAAAGATTGCAAGGATGCCAAAG 136
   |||
67 luThrIleGlnGlyValSerGluTyrIleSerPheIleThrGlyGlu 83
   |||
137 AATCCGTTACAGAGTGTCTGAGTCAGTTTATAGCTTCATAACATCTGAA 186
   |||
84 AlaAsnGluArgCysGlnArgGluGlnArgLysThrIleThrAlaGluAs 100
   |||
187 GCAAGTCAGCGCTCTCACCAGGAGAGCGGCAAGACCATCAACGGGAGGA 236
   |||
100 pValLeuTrpAlaMetSerArgLeuGlyPheAspTyrValGluProl 117
   |||
237 CATTCCTGTGCGCATGTCCACCTCTCGGCTTCGACAGCTACGTGGAGCCTC 286
   |||
117 euGlyAlaTyrLeuHisArgTyrArgGluPheGluGlyAspAlaArgGly 133
   |||
287 TGAACACTACTCTCCAGAAGTTCAGAGAGGCCCATGAAGGGAGAGAGGGC 336
   |||
134 ValGlyLeuValProGlyAlaAlaProSerArgGlyGly 146
   |||
337 ATTGGT.....GGGGCGGTGTCTGTCTACAGATGGA 366
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 25, 2001, 10:47:45 ; Search time 1233.41 Seconds
(without alignments)
14710.169 Million cell updates/sec

Title: US-09-435-054-1
Perfect score: 1173
Sequence: 1 ccacgcgtccgccaccacac.....tcaaaaaaaaaaaaaaaaaa 1173

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues
Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*

- 1: gb_bal:*
- 2: gb_ba2:*
- 3: gb_ba3:*
- 4: gb_in1:*
- 5: gb_in2:*
- 6: gb_in3:*
- 7: gb_om:*
- 8: gb_ov:*
- 9: gb_pat1:*
- 10: gb_pat2:*
- 11: gb_ph:*
- 12: gb_pl1:*
- 13: gb_pl2:*
- 14: gb_pl3:*
- 15: gb_pl4:*
- 16: em_bal:*
- 17: em_ba2:*
- 18: em_fun:*
- 19: em_htgo_hum:*
- 20: em_htgo_inv:*
- 21: em_htgo_rod:*
- 22: em_htg_hum1:*
- 23: em_htg_hum2:*
- 24: em_htg_hum3:*
- 25: em_htg_hum4:*
- 26: em_htg_hum5:*
- 27: em_htg_hum6:*
- 28: em_htg_hum7:*
- 29: em_htg_hum8:*
- 30: em_htg_inv1:*
- 31: em_htg_inv2:*
- 32: em_htg_other:*
- 33: em_htg_rod:*
- 34: em_hum1:*
- 35: em_hum2:*
- 36: em_hum3:*
- 37: em_hum4:*
- 38: em_hum5:*
- 39: em_hum6:*
- 40: em_hum7:*
- 41: em_in:*
- 42: em_om:*
- 43: em_or:*

44: em_ov:*

45: em_pat:*

46: em_ph:*

47: em_pl:*

48: em_ro:*

49: em_sts:*

50: em_sy:*

51: em_un:*

52: em_vi:*

53: gb_sts1:*

54: gb_sts2:*

55: gb_sts3:*

56: gb_sy:*

57: gb_un:*

58: gb_vil:*

59: gb_vil2:*

60: gb_htg1:*

61: gb_htg2:*

62: gb_htg3:*

63: gb_htg4:*

64: gb_htg5:*

65: gb_htg6:*

66: gb_htg7:*

67: gb_htg8:*

68: gb_htg9:*

69: gb_htg10:*

70: gb_htg11:*

71: gb_htg12:*

72: gb_htg13:*

73: gb_htg14:*

74: gb_htg15:*

75: gb_htg16:*

76: gb_htg17:*

77: gb_htg18:*

78: gb_htg19:*

79: gb_htg20:*

80: gb_htg21:*

81: gb_htg22:*

82: gb_htg23:*

83: gb_htg24:*

84: gb_htg25:*

85: gb_pr1:*

86: gb_pr2:*

87: gb_pr3:*

88: gb_pr4:*

89: gb_pr5:*

90: gb_pr6:*

91: gb_pr7:*

92: gb_pr8:*

93: gb_pr9:*

94: gb_rod:*

95: gb_rod2:*

96: gb_in4:*

97: gb_pr10:*

98: em_ba3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	199.8	17.0	141808	83	AP003266
2	199.8	17.0	179686	83	AP003246
3	198.6	16.9	80117	12	AB025628
4	179	15.3	826	12	AF036684
5	177.4	15.1	82875	12	AC013482
6	158.2	13.5	870	15	ZMNFYB
7	142.2	12.1	874	13	ATHAP3B
8	138.2	11.8	699	8	PMNFYB
					AP003266 Oryza sat
					AP003246 Oryza sat
					AB025628 Arabidops
					AF036684 Arabidops
					AC013482 Genomic s
					X59714 Z.mays mrna
					Y13724 Arabidops
					X59712 P.marinus m

TITLE Direct Submission
JOURNAL Submitted (19-FEB-2001) Takuji Sasaki, National Institute of
 Agrobiological Resources, Rice Genome Research Program, Kannondai
 2-1-2, Tsukuba, Ibaraki 305-8602, Japan
 (E-mail:tsasaki@abrr.affrc.go.jp, URL:http://rnp.dna.affrc.go.jp/,
 Tel:81-298-38-7441, Fax:81-298-38-7468)
COMMENT NOTE: It currently consists of 1 contigs. Gaps between the contigs
 are represented as runs of N. The order of the pieces is believed
 to be correct as given, however the sizes of the gaps between them
 are based on estimates that have provided by the submitter. This
 sequence will be replaced by the finished sequence as soon as it is
 available and the accession number will be preserved.
 * NOTE: This is a 'working draft' sequence.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.
FEATURES Location/Qualifiers
 source
 1..179686
 /organism="Oryza sativa"
 /cultivar="Nipponbare"
 /db_xref="taxon:4530"
 /chromosome="1"
 /clone="P0423A12"
BASE COUNT 51983 a 38309 c 38396 g 50748 t 250 others
ORIGIN
 Query Match 17.0% Score 199.8; DB 83; Length 179686;
 Best Local Similarity 65.7%; Pred. No. 3.2e-20;
 Matches 291; Conservative 0; Mismatches 152; Indels 0; Gaps 0;
 Qy 136 atggcggcgtctcagcagcagtcggcgccgagtcctccgagcagcagcgtctgctgc 195
 Db 28008 AGGGGGAGCGCGCGATCAGCGCGCGGAGATCATCAAGGAGCAGCAGGTCTGTCG 28067
 Qy 196 cgatcgcaacgtgacatcgatcgcgcggtgctgctgcgagcagcagcagcagtcg 255
 Db 28068 CAATCGGAGCGTGGCGGGAATCATGAAGCAGATCTCTCCCGCCCAAGCGCAAGATCTCCA 28127
 Qy 256 acgacgcaaggagacatccagagtgctgctcgagtgatcatcagcttcacgcgggg 315
 Db 28128 AGGAGGCCAAGGACGACGAGGAGTGCCTCTCGGAGTTCATCAGCTTCGACCGCGG 28187
 Qy 316 aggcacacgagcgtgcagcagcggagcagcagcagcagcagcagcagcagcagtcgt 375
 Db 28188 AGGCTTCGACAAAGTGCACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 28247
 Qy 376 gggccatgagcgcctcgcttcagcagcagcagcagcagcagcagcagcagcagcagc 435
 Db 28248 GGGCTTCGGCGCCCTCGGCTTCGACGACGACGACGACGACGACGACGACGACGACG 28307
 Qy 436 gctaccgaggttcgagggcgagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 495
 Db 28308 AGTACCGAGGTCGAGGGCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 28367
 Qy 496 cgagcggcgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 555
 Db 28368 GCG 28427
 Qy 556 ggcagctctcggagc 578
 Db 28428 GTCATTCGATGTTCAAGGCCATG 28450
RESULT 3
LOCUS AB025628 80117 bp DNA PLN 27-DEC-2000
DEFINITION Arabidopsis thaliana genomic DNA, chromosome 5, p1 clone:MNJ7.
ACCESSION AB025628 BA000015
VERSION AB025628.1 GI:4589434
KEYWORDS Arabidopsis thaliana (strain:Columbia) DNA, clone_lib:Mitsui P1
SOURCE clone:MNJ7.

ORGANISM

Arabidopsis thaliana
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
 Brassicales; Brassicaceae; Arabidopsi.
 1 (sites)

REFERENCE

AUTHORS

Kaneko,T., Katoh,T., Asamizu,E., Sato,S., Nakamura,Y., Kotani,H.
 and Tabata,S.

TITLE

Structural analysis of Arabidopsis thaliana chromosome 5. XI

JOURNAL

Unpublished

REFERENCE

2 (bases 1 to 80117)

AUTHORS

Nakamura,Y.

TITLE

Direct Submission

JOURNAL

Submitted (02-APR-1999) Yasukazu Nakamura, Kazusa DNA Research

Institute, Department of Plant Gene Research; 1532-3, Yana,

Kisarazu, Chiba 292-0812, Japan (E-mail:ynakamu@kazusa.or.jp,

Tel:81-438-52-3935, Fax:81-438-52-3934)

Address for correspondence: kaos@kazusa.or.jp

For the latest information on annotation of this clone, please see

http://www.kazusa.or.jp/kaos/cgi-bin/agd_graph.cgi?c=mnj7

Genes with similarity to proteins in the databases are described in

'product' or 'note' qualifiers. Genes that have no significant

protein similarity are described as 'unknown protein'.

The software programs used to predict genes include: Grail

(Informatics Group, Oak Ridge National Laboratory,

http://compbio.ornl.gov/Grail-1.3/),

GENSCAN (Chris Burge, MIT, http://CCR-081.mit.edu/GENSCAN.html),

NetGene2 (S.M. Hebsgaard, et al., CBS, Technical University of

Denmark, http://www.cbs.dtu.dk/services/NetGene2/) and

SplicePredictor (Volker Brendel, Stanford University,

http://gremli1.zool.iastate.edu/cgi-bin/sp.cgi).

Genes encoding tRNAs are predicted by tRNAscan-SE

(Sean Eddy, Washington University School of Medicine, St. Louis,

http://genome.wustl.edu/eddy/tRNAscan-SE/).

This sequence may not be the entire insert of this clone. It may be

shorter because we remove overlaps between neighboring submissions.

The 5' clone is MQL5 and the 3' clone is MGL1.

Location/Qualifiers

1..80117

/organism="Arabidopsis thaliana"

/strain="Columbia"

/db_xref="taxon:3702"

/chromosome="5"

/clone="MNJ7"

/clone_lib="Mitsui P1"

complement(12..108)

/note="CDS is reported in Acc# AB018117

gene_id:MQL5.29"

/number=2

/evidence=not_experimental

/product="DNA-binding protein-like"

complement(375..540)

/note="CDS is reported in Acc# AB018117

gene_id:MQL5.29"

/number=1

/evidence=not_experimental

/product="DNA-binding protein-like"

join(1123..1976,2116..2266,2344..2417,2508..2667,

2834..2932,3028..3114,3222..3365)

/note="gene_id:MNJ7.1"

/codon_start=1

/pseudo

/evidence=not_experimental

/product="formyltetrahydrofolate deformylase-like"

complement(join(3514..3755,3836..3893,3983..4032,

4110..4289,4380..4773,4852..5148))

/note="gene_id:MNJ7.2

pir||G71442

strong similarity to unknown protein"

/codon_start=1

/evidence=not_experimental

/protein_id="BAB09069.1"

/db_xref="GI:9758771"

/translation="MEGGFYSDWDSLSLFGSENPHELEGNVRSEIVSIPOQO

COMMENT

[illegible]

RESULT	6	
ZMNFYB		
LOCUS	870 bp	mRNA
DEFINITION	2..mays mRNA for CAAT-box DNA binding protein subunit B (NF-YB).	
ACCESSION	X59714	
VERSION	X59714.1	GI:22379
KEYWORDS	CAAT-box DNA binding protein.	
SOURCE	zea mays.	
ORGANISM	Zea mays	
REFERENCE	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Zea.	
AUTHORS	1 (bases 1 to 870)	
TITLE	Benoist,C.	
JOURNAL	Direct Submission	
REFERENCE	Submitted (21-JAN-1992) C. Benoist, L.G.M.E., Dept of Immunology, 11, Rue Humann, Strassbourg 67000, FRANCE	
AUTHORS	2 (bases 1 to 870)	
TITLE	Li,X.X., Mantovani,R., Hooft van Huijsduijnen,R., Andre,I., Benoist,C. and Mathis,D.	
JOURNAL	Evolutionary variation of the CCAAT-binding transcription factor NF-Y	
MEDLINE	Nucleic Acids Res. 20 (5), 1087-1091 (1992)	
REMARK	92195809	
FEATURES	Erratum:[{published erratum appears in Nucleic Acids Res 1992 Apr 11;20(7):1841}]	
SOURCE	Location/Qualifiers	
	1..870	
	/organism="Zea mays"	
	/db_xref="taxon:4577"	
	29..565	
	/codon_start=1	
	/product="CAAT-box DNA binding protein subunit B (NF-YB)"	
	/protein_id="CA4234.1"	
	/db_xref="GI:22380"	
	/db_xref="SWISS-PROT:P25209"	
	/translation="MAEPASPGGGSHESGPRGGGGSGVREQDRFLPTANISRI MKKAIIPANKIAKDAKETVQECSEFISITSEASDKCQREKRTINGDLLWAMATL GFDEYIEPLKVLQYRREMEGSGSKLTAKSGDSGSIKKDALGHVAGSSAABEGMQQAG NQGMGYMPQYHNGDIS"	
BASE COUNT	224 a	189 c 258 g 199 t
ORIGIN		

Query Match	13.5%	Score 158.2;	DB 15;	Length 870;
Best Local Similarity	68.2%;	Pred. No. 2.7e-13;		
Matches 234;	Conservative 0;	Mismatches 108;	Indels 1;	Gaps
<hr/>				
Qy	115	cggcgccgcccgcgcacaaatggcggcgtctcagcagcatgpgcgcgcgcgatcc	174	
Db	58	CGGCGGCGGGAGCCACGAGACGGGAGCCCCAGGGAGCGGAGCGTGGCAGCG-TCA	116	
Qy	175	gcgagcaggaccggcgtgatgccatcgcgcaactgataccgcatacatcgcgcgctgctgc	234	
Db	117	GGGAGCAGGCAGGTTCCTGCCCATCCGCCAATCATCTGCATCATGAAGAAGGCCATCC	176	
Qy	235	cggcgacgcgaagtatctcggaagcgcgcaaaggagacgatccaggagtgctgtctcgaggt	294	
Db	177	CGGTAAACGGGGAAGATCGCCAAGGAGCGCTAAGGAGACCGTGTCAGGAGTGCTCTCCGAGT	236	
Qy	295	acatcacgttcatacagggggagccaacgagcgggtccagcgggagcagcgaagacca	354	
Db	237	TCACTCTCTTATCATACGAGCAGCGAGTGACAAGTGCCAGAGGGAGAGCGGAAGACCACCA	296	
Qy	355	tcaocgcggaggacgtgctgtggccatgagcgacctcgagcttcgacgactagctcgagc	414	
Db	297	TCAATGCCGACGACTCTGCTGTGGCCATGGCCACGCTGGGGTTTGAAGACTACATTGTAAC	356	
Qy	415	cgctcggcgctaactccaccgcgtacogcgagttcagggcga	457	
Db	357	CCCTCAAAGGTGTACCTACAGAAGTACAGAGAGATGGAGGGGTGA	399	

RESULT	7	
LOCUS	ATHAP3B	
DEFINITION	Arabidopsis thaliana mRNA for Hap3b transcription factor.	12-SEP-1997
ACCESSION	Y13724	PLN
VERSION	Y13724.1	GI:2398528
KEYWORDS	hap3b gene; transcription factor.	
SOURCE	thale cress.	
ORGANISM	Arabidopsis thaliana	
REFERENCE	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.	
AUTHORS	1 (bases 1 to 874)	
TITLE	Edwards,D.	
JOURNAL	Direct Submission	
REFERENCE	Submitted (12-JUN-1997) Edwards D., Plant Sciences, University of Cambridge, Downing Street, Cambridge CB2 3EA UK	
AUTHORS	2 (bases 1 to 874)	
TITLE	Edwards,D., Smith,A.G. and Murray,J.A.	
JOURNAL	Isolation and characterisation of CCAAT box binding proteins from higher plants	
FEATURES	Unpublished	
SOURCE	Location/Qualifiers	
gene	1..874	
CDS	/organism="Arabidopsis thaliana"	
	/db_xref="taxon:3702"	
	1..565	
	/gene="hap3b"	
	<1..565	
	/gene="hap3b"	
	/codon_start=2	
	/product="transcription factor"	
	/protein_id="CAA74052.1"	
	/db_xref="GI:2398529"	
	/db_xref="SPTREMBL:O23634"	
	/translation="RRDRSGGGQNGNQSSLSPREQDFLPFIANVSRIMKKAALPA NAKISKDAKMEQCVSEFISRVTEASDKCKQKRRKTINDGILLWMTTIGPDEVYE PLKVIYLRFRFEGREITGLRPTGGVEGHRDAVGGGGGYGGGGMOYHOHHOFL HQQNHMYGATGGDSGGGAAASGRTRT"	
BASE COUNT	263 a	155 c 244 g 212 t
ORIGIN		

Query Match 12.18; Score 142.2; DB 13; Length 874;

Qy 169 cgaatccgcgagcagaccgcgctgatacccatcgcaacgtgtatccgcatactgcggcgcg 228
/gene="AT4g14420"
complement(5773..5864)
intron
intron

AUTHORS

JOURNAL

COMMENT

Lin, X.
 Direct Submission
 Submitted (09-MAR-2000) The Institute for Genomic Research, 9712
 Medical Center Dr., Rockville, MD 20850, USA
 On Dec 17, 1999 this sequence version replaced gi:3738275.
 The sequence and annotation of chromosome 2 were merged from those
 of the individual clones on this chromosome after removing
 overlaps. For detailed information, please see the TIGR web site
 (<http://www.tigr.org/tdb/at/at.html>).

Genes were identified by a combination of three methods: Gene
 prediction programs including GRAIL
 (<ftp://arthur.epm.ornl.gov/pub/xgrail/>), Genefinder (Phil Green,
 University of Washington), Genscan (Chris Burge,
<http://gnomic.stanford.edu/GENSCANW.html>), and NetPlantGene
 (<http://www.cbs.dtu.dk/services/NetGene2/>), searches of the
 complete sequence against a peptide database and plant EST
 databases at TIGR, and manual curations based on those analyses.
 Annotated genes are named to indicate the level of evidence for
 their annotation. Genes with similarity to other proteins are named
 after the database hits. Genes without significant peptide
 similarity but with EST similarity are named as 'unknown' proteins.
 Genes without protein or EST similarity, that are predicted by two
 or more gene prediction programs over most of their length are
 annotated as 'hypothetical' proteins. Genes encoding tRNAs are
 predicted by tRNAscan-SE (Sean Eddy,
<http://genome.wustl.edu/eddy/tRNAscan-SE/>). Simple repeats were
 identified by RepeatMasker (Arian Smit,
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>). Genes are
 numbered from the top to bottom of the chromosome.

We thank the CSHL/WashU/ABI consortium for sequencing BAC clones
 F6P23, F5J6, T17A5, and T13L16, the ESSA group for sequencing clone
 F13D4, and Scott Jackson, Jining Jiang, Klaus Meyer, Eric Richards
 and Satoshi Tabata for helpful assistance. In addition, we would
 like to thank the TIGR Bioinformatics Department, especially Lixin
 Zhou, Hanif Khalak, Michael E. Heaney, Lily Fu, Feng Liang, Jeremy
 Peterson, Michael Holmes, and Delwood Richardson for software and
 database support.

This work was supported by the National Science Foundation,
 Department of Energy and the US Department of Agriculture.

Address all correspondence to: at@tigr.org.

FEATURES

source

Location/Qualifiers
 1. 107377
 /organism="Arabidopsis thaliana"
 /cultivar="Columbia"
 /db_xref="taxon:3702"
 /chromosome="II"

misc_feature

complement(<1..821)
 /note="Sequence from clone T30B22"

join(<291..348,504..1024,1120..1549,2112..2233,
 2332..>2508)

/gene="At2g47620"

<291..>2508

/gene="At2g47620"

/note="T30B22.7"

join(291..348,504..1024,1120..1549,2112..2233,2332..2508)

/gene="At2g47620"

/codon_start=1

/product="putative SWI/SNF family transcription activator"

/protein_id="AAC63618.2"

/db_xref="GI:6598454"

/translation="MEATDPSAEIYLTIPAOSSWFLWDDIHIERREFAEPTESSI
 TRPKYKVEDIINKFREDTERRLTFTSVKFLYGVNLLQKVFLEKGLNFS
 SSLKNDHLSDVLAECQTPAGIRVATPNSLRIPAPLVEERVETGIKVPPLTS
 YSDVFDLKPDLVCAHCGCRDSPFYOHKNICVCEKFCNNGYENNTADDFK
 LIGNSAAVTEEEIILLLESVLKHGDDWELISQSVSTRKLDICSKLIEPFGFLM
 GSASGRNLNEDTEOVQDQGEHETETREKEDRVNEDPEPPAKRKVALISE
 GDTLRIRASVATGAAAOAKILADQEEEMEQLAATVIEOQLKLSKLFLLDLE
 SIMDEEKVTEGKETIQRVSVLOCAFRSGITRKWDHTYVK"

misc_feature

822..107176

repeat_region

misc_feature

/note="Sequence from clone F17A22"
 complement(1028..1102)
 /rpt_family="(CAAAA)n"
 1216..1715

/gene="At2g47620"

/note="molecular marker mi79a"

join(<3826..4057,4175..4305,4398..4535,4614..4712,
 4788..>5168)

/gene="At2g47630"

<3826..>5168

/gene="At2g47630"

/note="F17A22.2"

join(3826..4057,4175..4305,4398..4535,4614..4712,
 4788..5168)

/gene="At2g47630"

/codon_start=1

/product="putative phospholipase"

/protein_id="AAC63619.1"

/db_xref="GI:3738277"

/translation="MASLVVEKIYSEVDTCLKNLKHVLLECGIRLASAGYAVFGM
 DYEGHGRSMGRSRYIKKFNANVYNDYDYSICAEYMDKGRFLYGSMGAVTLLL
 HKKDLFWNAILVAPMCKISIKRAPHFIVINLTVRVEIIPKMKIVPTKVIDRAFK
 DLVKEEVNKKLIYQDPRKTALEMLRTSMNLEDTHIETMPFVJHGSDVTPD
 EYSKALYKASTRDKTLKLYPGMHALTSGEPCNVDLVFADIIINLDRTPADPSLT
 VPIRVGNTTSVQRTVTYGVSGNHRPRKRPFFNLLCGINRGLVPRSTV"

join(5465..5512,5607..5654,5762..5786,6120..6194,
 6385..6489,6572..6844)

/gene="At2g47640"

5465..6844

/gene="At2g47640"

/note="F17A22.3"

join(5650..5654,5762..5786,6120..6194,6385..6489,
 6572..6688)

/gene="At2g47640"

/codon_start=1

/product="putative small nuclear ribonucleoprotein D2"

/protein_id="AAC63620.1"

/db_xref="GI:3738278"

/translation="MSKPMEDTNGKTEEFNTGTLVLMKSVKNNQTVLINCNNR
 KLLGRVAFDRHCNMVLENVREMTVEVPTGKGGKALPVNRDRFISKMFRLGDSVII
 VLNRPK"

complement(join(6904..7235,7321..7413,7965..8039,
 8128..8211,8290..8369,8740..8887,8998..9714))

/gene="At2g47650"

complement(6904..9714)

/gene="At2g47650"

/note="F17A22.4"

complement(join(7008..7235,7321..7413,7965..8039,
 8128..8211,8290..8369,8740..8887,8998..9621))

/gene="At2g47650"

/note="The first intron is U12-dependent (AT/AC class)"

/codon_start=1

/product="putative nucleotide-sugar dehydratase"

/protein_id="AAC63621.1"

/db_xref="GI:3738279"

/translation="MASELTNRHREIEQEAESYYPKPKPWFAIRPIRYMLRQRL
 VFVLGIATLTGFTIFSKSNHOPIDYDVPDLSYGMRSSESYLPATIHKKPSIEYM
 SFRISAGGKIPGLKARKLVVVVTGGAGVSHLDRLMARGDNVIVVDNFTGRKEN
 VNHNNFNMIRHDVVEPILLEDVQIYHLACPASPHYKFNPKTKITNVYGLNM
 LGLARKVGARELLTSTSEYGDLPQHPQVETWGNVNPIGVRSYDEGKRTAETLMD
 YHRGANVEVRIARTIENTYGRMCIIDDGRVSNVFAQALKEPLTVYDGKQTRSFQFV
 SDLVGLMLRLEGEHGPENLGNPGFTMLELAKVQVETIDPNAKIEFRPNTDDPHK
 RKPDITTKAKELILGWEPKVALRQGLPLMYKFRQVRQGVQKQDSSTSSSTE"

7061..9447

/note="molecular marker 90J19T7"

8690..8736

/rpt_family="(GAA)n"

join(<11209..11311,11413..>11417)

/gene="At2g47660"

<11209..>11417

/gene="At2g47660"

/note="F17A22.5; predicted by genscan"

join(11209..11311,11413..11417)

misc_feature

7061..9447

repeat_region

8690..8736

misc_feature

8690..8736

misc_feature

8690..8736

misc_feature

8690..8736

misc_feature

8690..8736

misc_feature

8690..8736

misc_feature

8690..8736

misc_feature

8690..8736

misc_feature

8690..8736

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 266340)

Adams, M.D., Celniker, S.E., Holt, R.A., Evans, C.A., Gocayne, J.D., Amanatides, P.G., Scherer, S.E., Li, P.W., Hoskins, R.A., Galle, R.F., George, R.A., Lewis, S.E., Richards, S., Ashburner, M., Henderson, S.N., Sutton, G.G., Wortman, J.R., Yandell, M.D., Zhang, Q., Chen, L.X., Brandon, R.C., Rogers, Y.H., Blaziej, R.G., Champ, M., Pfeiffer, B.D., Wan, K.H., Doyle, C., Baxter, E.G., Helt, G., Nelson, C.R., Gabor, Miklos, G.L., Abri, J.F., Agbayani, A., An, H.J., Andrews-Pfannkoch, C., Baldwin, D., Ballew, R.M., Basu, A., Baxendale, J., Bayraktaroglu, L., Beasley, E.M., Beeson, K.Y., Benos, P.V., Berman, B.P., Bhandari, D., Bolshakov, S., Borkov, D., Botchan, M.R., Bouck, J., Brockstein, P., Brottier, P., Burtis, K.C., Busam, D.A., Butler, H., Cadieux, E., Center, A., Chandra, I., Cherry, J.M., Cawley, S., Dahlke, C., Davenport, L.B., Davies, P., de Pablos, B., Delcher, A., Deng, Z., Mays, A.R.D., Dew, I., Dietz, S.M., Dodson, K., Doup, L.E., Downes, M., Dugan-Rocha, S., Dunkov, B.C., Dunn, P., Durbin, K.J., Evangelista, C.C., Ferraz, C., Ferreira, S., Fleischmann, W., Folsler, C., Gabrielian, A.E., Garq, N.S.,

REFERENCE
AUTHORS

```

Gelbart,W.M., Glasser,K., Glodek,A., Cong,F., Gorrell,J.H., Gu,Z.,
Guan,P., Harris,M., Harris,N.L., Harvey,D., Helman,T.J.,
Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J.,
Wei,M.H., Ibegwam,C., Jalali,M., Kalush,F., Karpen,G.H., Ke,Z.,
Kennison,J.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C.,
Kravitz,S., Kulp,D., Lai,Z., Lasko,P., Lei,Y., Levitsky,A.A.,
Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Mattei,B., McIntosh,T.C.,
McLeod,M.P., McPherson,D., Merkulov,G., Milshina,N.V., Mobarri,C.,
Morris,J., Moshref,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L.,
Muzny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K.,
Nusskern,D.R., Pacleb,J.M., Palazzolo,M., Pittman,G.S., Pan,S.,
Pollard,J., Puri,V., Reese,M.G., Reinert,K., Remington,K.,
Saunders,R.D., Scheeler,F., Shen,H., Shue,B.C., Siden-Kiamos,I.,
Simpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C.,
Stapleton,M., Strong,R., Sun,E., Swirskas,R., Tector,C., Turner,R.,
Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wasserman,D.A.,
Weinstock,G.M., Weissbach,J., Williams,S.M., Woodage,T.,
Worley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,J., Yeh,R.F.,
Zaveri,J.S., Zhan,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H.,
Zhong,F.N., Zhong,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O.,
Gibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,J.C.
The genome sequence of Drosophila melanogaster
Science 287 (5461), 2185-2195 (2000)
20196006
2 (bases 1 to 266340)
Adams,M.D., Celniker,S.E., Gibbs,R.A., Rubin,G.M. and Venter,C.J.
Direct Submission
Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
FEATURES
source
1. 266340
Location/Qualifiers
organism="Drosophila melanogaster"
db_xref="taxon:7227"
chromosome="2L"
complement(join(542..567,703..1027,1121..1407,1597..2899,
2958..3179,3470..3591,3725..>3906))
/gene="CG10462"
/product="CT29376"
/db_xref="FLYBASE:FBan0010462"
/db_xref="FLYBASE:FBgn0032815"
complement(<542..>3906)
/gene="CG10462"
/map="38A2-38A2"
/db_xref="FLYBASE:FBan0010462"
/db_xref="FLYBASE:FBgn0032815"
complement(join(542..567,703..1027,1121..1407,1597..2899,
2958..3179,3470..3591,3725..3794))
/gene="CG10462"
/note="CG10462 gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0010462"
/db_xref="FLYBASE:FBgn0032815"
/protein_id="AAF53838.1"
/db_xref="GI:7298622"
/translation="MEDIKMEPQDEVTSSIRLSACDTLKILDRGSMITVPIRD
RLRSDERLFTASLEGLHQIIFASEEDNPKNHCDCIQKFSQDIISSHRVHHEMN
LHFCRLGCGIWDLTLEIHYRQHPQGRVLCRCYCDWFANSADOLASHMOKHIYV
YFVCSCTRHFESKQSLNLRKQSDFCILLDFKGSXSKPELVAVSTVPRESNCY
VQIKEEPLNAEARMLEDDNPPLNKLQDEEMPDSYRIKEQSGNSVWTAIHTPL
NKLRLPALATIKOMHLKRFPAINKKSSIIITTFSAVKRNGLLTSNVELNKS
NHSVTDADNIPISNPFSSIAANILKVLPSNCMIKLPNTKIFKTPGQAPAPK
DSPTVLTSGTNSNFNIKPIKTSISTVSGMGNPHASAKTVQPTQTHSRSSCFLVDSF
NNSKSRPESMKIINEFRNIRSIETLPLPGTVLQSKVMDESTSSPMDAPKPELVM
LNPSALMARELEIQPLYRPMWTCPLCQRMFEKHCAPRAHLTNKHLDTBEKNLVK
VLMPSKTSALSNVPYNKGLSPISINIEKPKSKKNQSQCTCLAVFTFGAL
RHKSHHGELPQCSYCDKFRTPGQVRVHRRHRTGKEPKKICSLDTHRETLIS
HLSRHGMKRYKCYGCDKRVFVYVSLGRAHRLRPDTCGVKFTARAHPKRVIRGEV
VFQYHPENHGVLRSDEPLNLSQRDQTNSTPETAIWMRGITPE"
join(<4094..4166,4231..>4709)
/gene="CG10447"
/product="CT29328"
/db_xref="FLYBASE:FBan0010447"
/db_xref="FLYBASE:FBgn0032816"

```

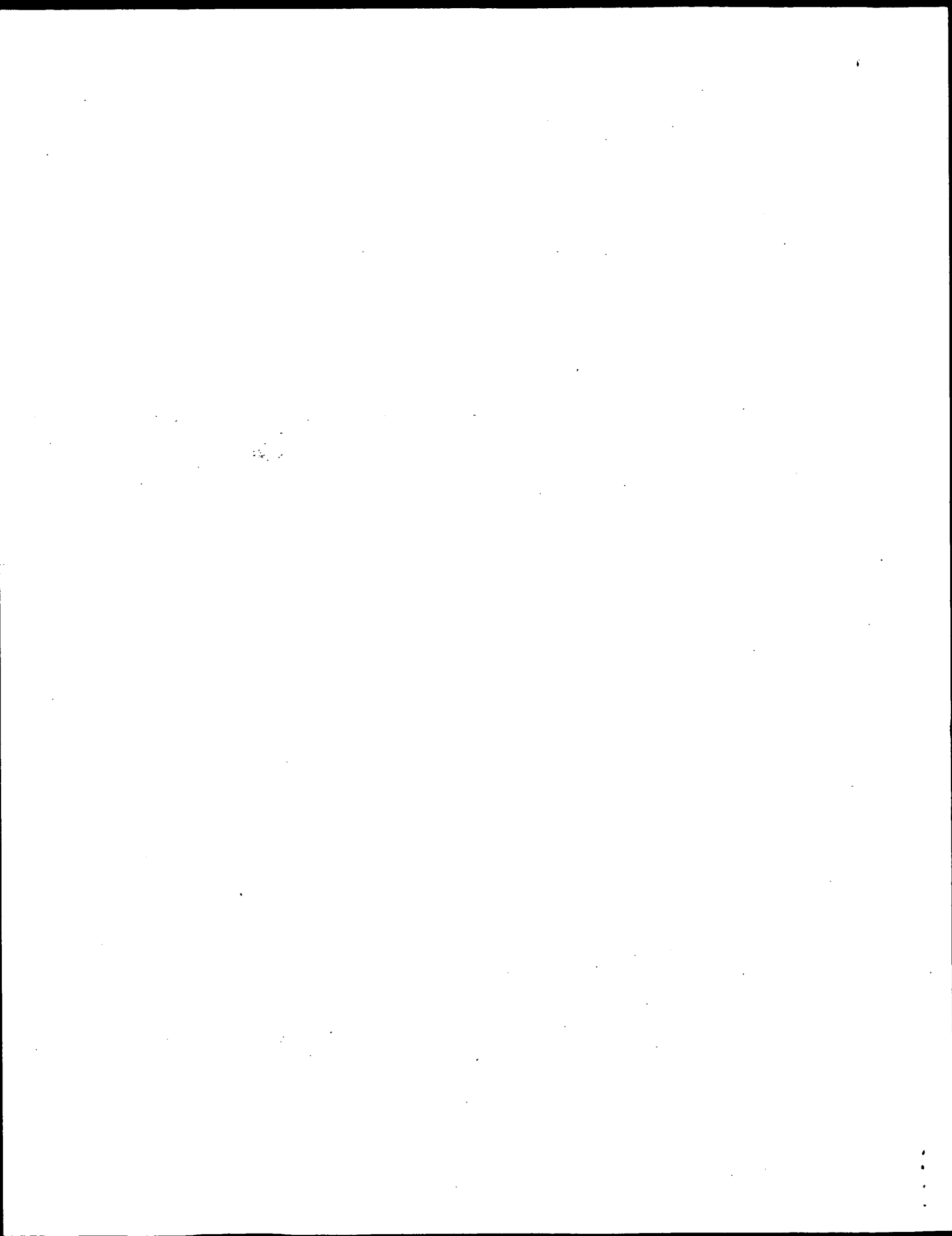
gene

CDS

mRNA

gene

CDS



GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 25, 2001, 09:19:30 ; Search time 12.26 Seconds
(without alignments)
466.894 Million cell updates/sec

Title: US-09-435-054-2
Perfect score: 1514
Sequence: 1 MDSSFLPAAGENSAGG.....ASHTPOGSGGLEHPPHAYK 278

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 197339 seqs, 20590346 residues

Total number of hits satisfying chosen parameters: 197339

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: /cgn2_6/ptodata/2/iaa/5A.COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B.COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A.COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B.COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PTCUS.COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	494	32.6	208	4	US-09-103-478-2
2	403	26.6	90	4	US-09-103-478-19
3	341	22.5	90	4	US-09-103-478-20
4	305	20.1	90	4	US-09-103-478-22
5	301	19.9	90	4	US-09-103-478-21
6	301	19.9	90	4	US-09-103-478-24
7	301	19.9	90	4	US-09-103-478-25
8	298	19.7	90	4	US-09-103-478-26
9	295.5	19.5	144	4	US-09-319-989-4
10	270	17.8	85	4	US-09-103-478-28
11	269	17.8	85	4	US-09-103-478-27
12	262	17.3	90	4	US-09-103-478-29
13	208	13.7	57	4	US-09-103-478-23
14	160	10.6	156	1	US-09-681-812-7
15	147	9.7	1185	4	US-09-041-886-23
16	140.5	9.3	420	2	US-08-845-998-8
17	140.5	9.3	420	2	US-08-206-537-8
18	129.5	8.6	918	4	US-09-041-886-11
19	120	7.9	1958	1	US-07-945-283-2
20	118	7.8	738	3	US-08-864-038A-3
21	118	7.8	907	2	US-09-010-928B-4
22	114	7.5	410	4	US-09-431-573-5
23	112	7.4	553	3	US-09-083-351-2
24	112	7.4	553	4	US-09-083-352-2
25	112	7.4	745	2	US-09-010-928B-28
26	112	7.4	870	2	US-09-010-928B-2
27	109	7.2	410	4	US-09-431-573-4

28	107	7.1	410	1	US-08-123-343A-5	Sequence 5, Appli
29	107	7.1	410	1	US-08-123-343A-7	Sequence 7, Appli
30	106	7.0	340	5	PCT-US96-02331-13	Sequence 13, Appl
31	106	7.0	842	5	PCT-US96-02331-15	Sequence 15, Appl
32	105.5	7.0	2100	1	US-08-477-509B-80	Sequence 80, Appl
33	105.5	7.0	2100	3	US-08-482-085B-80	Sequence 80, Appl
34	105.5	7.0	2107	1	US-08-175-155-45	Sequence 45, Appl
35	105.5	7.0	2107	2	US-08-707-237A-51	Sequence 51, Appl
36	104.5	6.9	330	2	US-08-712-948-2	Sequence 2, Appli
37	104	6.9	333	2	US-08-712-948-1	Sequence 1, Appli
38	103.5	6.8	1057	3	US-08-931-820-1	Sequence 1, Appli
39	103.5	6.8	1341	3	US-08-963-825-18	Sequence 18, Appl
40	103	6.8	671	3	US-09-121-321-16	Sequence 16, Appl
41	103	6.8	671	4	US-08-933-803A-16	Sequence 16, Appl
42	103	6.8	1038	1	US-07-609-716-36	Sequence 36, Appl
43	103	6.8	1038	4	US-08-475-411A-36	Sequence 36, Appl
44	103	6.8	1038	4	US-08-478-029A-36	Sequence 36, Appl
45	102.5	6.8	595	1	US-08-425-069-4	Sequence 4, Appli

ALIGNMENTS

RESULT 1
US-09-103-478-2
; Sequence 2, Application US/09103478
; Patent No. 6235975
; GENERAL INFORMATION:
; APPLICANT: Harada, John
; APPLICANT: Lotan, Tamar
; APPLICANT: Ohto, Masa-aki
; APPLICANT: Goldberg, Robert B.
; APPLICANT: Fischer, Robert L.
; TITLE OF INVENTION: LEAFY COTYLEDON1 Genes and Their Uses
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/103,478
; FILING DATE: 24-JUN-1998
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 09/026,221
; FILING DATE: 19-FEB-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/804,534
; FILING DATE: 21-FEB-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Einhorn, Gregory P.
; REGISTRATION NUMBER: 38,440
; REFERENCE/DOCKET NUMBER: 023070-077611US
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 208 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-103-478-2

```
Query Match      22.5%; Score 341; DB 4; Length 90;
Best Local Similarity 65.6%; Pred. NO. 1.le-24;
Matches 59; Conservative 19; Mismatches 12; Indels 0; Gaps 0
```


Db 1 REQDRLPIANIRIMKKAIPANGKIAKAKETVQECVSEFISFITSEASDKQCKRKT 60
QY 96 ITAEDVLWMSRLGFDYVEPLGAYLHYR 125
Db 61 INGDLLWAMATLGFEDYIEPLKYLQYR 90

RESULT 4
US-09-103-478-22
; Sequence 22, Application US/09103478
; Patent No. 6235975
; GENERAL INFORMATION:
; APPLICANT: Harada, John
; APPLICANT: Lotan, Tamar
; APPLICANT: Ohto, Masa-aki
; APPLICANT: Goldberg, Robert B.
; APPLICANT: Fischer, Robert L.
; TITLE OF INVENTION: LEAFY COTYLEDON1 Genes and Their Uses
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/103,478
; FILING DATE: 24-JUN-1998
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 09/026,221
; FILING DATE: 19-FEB-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/804,534
; FILING DATE: 21-FEB-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Einhorn, Gregory P.
; REGISTRATION NUMBER: 38,440
; REFERENCE/DOCKET NUMBER: 023070-077611US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 90 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-103-478-22

Query Match 20.1%; Score 305; DB 4; Length 90;
Best Local Similarity 64.4%; Pred. No. 2.3e-21;
Matches 58; Conservative 13; Mismatches 19; Indels 0; Gaps 0;

QY 36 REQDRLPIANIRIMRRVLPAHAKISDDAKETIQECVSEYISFITGEANERCOREQRT 95
Db 1 REQDIYLPANVARIMKTSIPSSKIAKAKECVQECVSEFISFITSEASERCHQCKRKT 60
QY 96 ITAEDVLWMSRLGFDYVEPLGAYLHYR 125
Db 61 INGDILFAMSTLGFQSYVEPLKYLQYR 90

RESULT 5
US-09-103-478-21

; Sequence 21, Application US/09103478
; Patent No. 6235975
; GENERAL INFORMATION:
; APPLICANT: Harada, John
; APPLICANT: Lotan, Tamar
; APPLICANT: Ohto, Masa-aki
; APPLICANT: Goldberg, Robert B.
; APPLICANT: Fischer, Robert L.
; TITLE OF INVENTION: LEAFY COTYLEDON1 Genes and Their Uses
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/103,478
; FILING DATE: 24-JUN-1998
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 09/026,221
; FILING DATE: 19-FEB-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/804,534
; FILING DATE: 21-FEB-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Einhorn, Gregory P.
; REGISTRATION NUMBER: 38,440
; REFERENCE/DOCKET NUMBER: 023070-077611US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 90 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-103-478-21

Query Match 19.9%; Score 301; DB 4; Length 90;
Best Local Similarity 63.3%; Pred. No. 5.5e-21;
Matches 57; Conservative 13; Mismatches 20; Indels 0; Gaps 0;

QY 36 REQDRLPIANIRIMRRVLPAHAKISDDAKETIQECVSEYISFITGEANERCOREQRT 95
Db 1 REQDIYLPANVARIMKNAIPQTGKIAKAKECVQECVSEFISFITSEASERCHQCKRKT 60
QY 96 ITAEDVLWMSRLGFDYVEPLGAYLHYR 125
Db 61 INGDILFAMSTLGFQSYVEPLKYLQYR 90

RESULT 6
US-09-103-478-24
; Sequence 24, Application US/09103478
; Patent No. 6235975
; GENERAL INFORMATION:
; APPLICANT: Harada, John
; APPLICANT: Lotan, Tamar
; APPLICANT: Ohto, Masa-aki
; APPLICANT: Goldberg, Robert B.
; APPLICANT: Fischer, Robert L.
; TITLE OF INVENTION: LEAFY COTYLEDON1 Genes and Their Uses

us-09-435-054-2.ra1

Fri Oct 26 13:40:50 2001

NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/103,478

FILING DATE: 24-JUN-1998

CLASSIFICATION: 800

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 09/026,221

FILING DATE: 19-FEB-1998

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/804,534

FILING DATE: 21-FEB-1997

ATTORNEY/AGENT INFORMATION:

NAME: Einhorn, Gregory P.

REGISTRATION NUMBER: 38,440

REFERENCE/DOCKET NUMBER: 023070-077611US

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 576-0200

TELEFAX: (415) 576-0300

INFORMATION FOR SEQ ID NO: 24:

SEQUENCE CHARACTERISTICS:

LENGTH: 90 amino acids

TYPE: amino acid

STRANDEDNESS:

TOPOLOGY: linear

MOLECULE TYPE: peptide

US-09-103-478-24

Query Match 19.9%; Score 301; DB 4; Length 90;
Best Local Similarity 63.3%; Pred. No. 5.5e-21;
Matches 57; Conservative 13; Mismatches 20; Indels 0; Gaps 0;

QY 36 REQRLMPANVIRIMRVLPAAKISDDAKETIQCVSEYISFITGEANERCOREQKRT 95
Db 1 REQDIYLPANVARIMKNAIPQTGKIADKCEVCVSEFISFITSEASERCHQEKRT 60
QY 96 ITAEDVLWMSRLGFDYVEPLGAYLHRYR 125
Db 61 INGEDIILFAMSTLGFQSYVEPLKLYLQKFR 90

RESULT 7
US-09-103-478-25
Sequence 25, Application US/09103478
Patent No. 6235975
GENERAL INFORMATION:
APPLICANT: Harada, John
APPLICANT: Lotan, Tamar
APPLICANT: Ohto, Masa-aki
APPLICANT: Goldberg, Robert B.
APPLICANT: Fischer, Robert L.
TITLE OF INVENTION: LEAFY COTYLEDON1 Genes and Their Uses
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/103,478
FILING DATE: 24-JUN-1998
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 09/026,221
FILING DATE: 19-FEB-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/804,534
FILING DATE: 21-FEB-1997
ATTORNEY/AGENT INFORMATION:
NAME: Einhorn, Gregory P.
REGISTRATION NUMBER: 38,440
REFERENCE/DOCKET NUMBER: 023070-077611US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 25:
SEQUENCE CHARACTERISTICS:
LENGTH: 90 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-103-478-25

Query Match 19.9%; Score 301; DB 4; Length 90;
Best Local Similarity 63.3%; Pred. No. 5.5e-21;
Matches 57; Conservative 13; Mismatches 20; Indels 0; Gaps 0;

QY 36 REQRLMPANVIRIMRVLPAAKISDDAKETIQCVSEYISFITGEANERCOREQKRT 95
Db 1 REQDIYLPANVARIMKNAIPQTGKIADKCEVCVSEFISFITSEASERCHQEKRT 60
QY 96 ITAEDVLWMSRLGFDYVEPLGAYLHRYR 125
Db 61 INGEDIILFAMSTLGFQSYVEPLKLYLQKFR 90
RESULT 8
US-09-103-478-26
Sequence 26, Application US/09103478
Patent No. 6235975
GENERAL INFORMATION:
APPLICANT: Harada, John
APPLICANT: Lotan, Tamar
APPLICANT: Ohto, Masa-aki
APPLICANT: Goldberg, Robert B.
APPLICANT: Fischer, Robert L.
TITLE OF INVENTION: LEAFY COTYLEDON1 Genes and Their Uses
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/103,478
FILING DATE: 24-JUN-1998
CLASSIFICATION: 800
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 09/026,221
FILING DATE: 19-FEB-1998
PRIOR APPLICATION DATA:
APPLICANT: Lotan, Tamar
FILING DATE: 21-FEB-1997
ATTORNEY/AGENT INFORMATION:
NAME: Einhorn, Gregory P.
REGISTRATION NUMBER: 38,440
REFERENCE/DOCKET NUMBER: 023070-077611US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 26:
SEQUENCE CHARACTERISTICS:
LENGTH: 90 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-103-478-26

Query Match 19.7%; Score 298; DB 4; Length 90;
Best Local Similarity 61.1%; Pred. No. 1e-20;
Matches 55; Conservative 19; Mismatches 16; Indels 0; Gaps 0;
QY 36 REQDRLMPIANVIRMRVLPAAHAKISDDAKETIQECVSEYISFITGEANERCQREKRT 95
Db 1 KEQDWLPPIANVARIMKLALENPAKIAKEAKECMQECVSEYISFITSEASEKQCEKRT 60
QY 96 ITAEDVLWMSRLGFDYVEPLGAYLHRYR 125
Db 61 VNGEDILFAMTSLGFENYAEALKIYLSKYR 90

RESULT 9
US-09-319-989-4
Sequence 4, Application US/09319989
Patent No. 6190914
GENERAL INFORMATION:
APPLICANT: Grivell, Leslie A.
APPLICANT: Teixeira de Mattos, Maarten J.
TITLE OF INVENTION: METHODS FOR MODULATING METABOLIC PATHWAYS OF
TITLE OF INVENTION: MICRO-ORGANISMS AND MICRO-ORGANISMS OBTAINABLE BY SAID
TITLE OF INVENTION: METHODS
FILE REFERENCE: 24615-20123.00
CURRENT APPLICATION NUMBER: US/09/319,989
EARLIER FILING DATE: 1999-06-14
EARLIER APPLICATION NUMBER: PCT/NL97/00688
EARLIER FILING DATE: 1997-12-12
EARLIER APPLICATION NUMBER: BPO 96203520
NUMBER OF SEQ ID NOS: 9
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 4
LENGTH: 144
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: HAP3
US-09-319-989-4

Query Match 19.5%; Score 295.5; DB 4; Length 144;
Best Local Similarity 50.4%; Pred. No. 3.2e-20;
Matches 60; Conservative 21; Mismatches 33; Indels 5; Gaps 2;
QY 8 PAAGANGSAAGANGGAQQAAPAREQDRLMPIANVIRMRVLPAAHAKISDDAKE 67
Db 13 PEDTQENG--GNASSGSLQQ--ISTREQDWLPINNVARLMKNTLPPSAKYSKDAKE 67
QY 68 TIQECVSEYISFITGEANERCQREKRTITAEDVLWMSRLGFDYVEPLGAYLHRYRE 126

Db 68 CMQECVSEYISFITSEASDRCAADKRRKTINGEDILISLHALGFENYAEVLKIYLAQYRQ 126
RESULT 10
US-09-103-478-28
Sequence 28, Application US/09103478
Patent No. 6235975
GENERAL INFORMATION:
APPLICANT: Harada, John
APPLICANT: Lotan, Tamar
APPLICANT: Ohto, Masa-aki
APPLICANT: Goldberg, Robert B.
APPLICANT: Fischer, Robert L.
TITLE OF INVENTION: LEAFY COTYLEDON1 Genes and Their Uses
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/103,478
FILING DATE: 24-JUN-1998
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 09/026,221
FILING DATE: 19-FEB-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/804,534
FILING DATE: 21-FEB-1997
ATTORNEY/AGENT INFORMATION:
NAME: Einhorn, Gregory P.
REGISTRATION NUMBER: 38,440
REFERENCE/DOCKET NUMBER: 023070-077611US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 28:
SEQUENCE CHARACTERISTICS:
LENGTH: 90 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-103-478-28

Query Match 17.8%; Score 270; DB 4; Length 90;
Best Local Similarity 56.7%; Pred. No. 4e-18;
Matches 51; Conservative 16; Mismatches 23; Indels 0; Gaps 0;
QY 36 REQDRLMPIANVIRMRVLPAAHAKISDDAKETIQECVSEYISFITGEANERCQREKRT 95
Db 1 REQDWLPINNVARLMKNTLPPSAKYSKDAKECMQECVSEYISFITSEASDRCAADKRT 60
QY 96 ITAEDVLWMSRLGFDYVEPLGAYLHRYR 125
Db 61 INGEDILISLHALGFENYAEVLKIYLAQYR 90

RESULT 11
US-09-103-478-27
Sequence 27, Application US/09103478
Patent No. 6235975
GENERAL INFORMATION:

APPLICANT: Harada, John
APPLICANT: Lotan, Tamar
APPLICANT: Ohto, Masa-aki
APPLICANT: Goldberg, Robert B.
APPLICANT: Fischer, Robert L.
TITLE OF INVENTION: LEAFY COTYLEDON1 Genes and Their Uses
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/103,478
FILING DATE: 24-JUN-1998
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 09/026,221
FILING DATE: 19-FEB-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/804,534
FILING DATE: 21-FEB-1997
ATTORNEY/AGENT INFORMATION:
NAME: Einhorn, Gregory P.
REGISTRATION NUMBER: 38,440
REFERENCE/DOCKET NUMBER: 023070-077611US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 29:
SEQUENCE CHARACTERISTICS:
LENGTH: 90 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-103-478-29

Query Match 17.8%; Score 262; DB 4; Length 90;
Best Local Similarity 56.2%; Pred. No. 2.2e-17;
Matches 50; Conservative 15; Mismatches 24; Indels 0; Gaps 0;

QY 37 EODRLPIANVIRMRVLPAAHAKISDDAKETIQCVSEYISFITGEANERQCQRKRTI 96
Db 2 EQDRWLPINNVARLMKNTLPATTKVSKDAKCMQCVSEFISFVTSEACDRCTSGKRRKI 61
QY 97 TAEDVLWAMSRIGFDDYVEPLGAYLHYR 125
Db 62 NGEDILLSLHALGFENYAEVLKYLAKYR 90

RESULT 13
US-09-103-478-23
Sequence 23, Application US/09103478
Patent No. 6235975
GENERAL INFORMATION:
APPLICANT: Harada, John
APPLICANT: Lotan, Tamar
APPLICANT: Ohto, Masa-aki
APPLICANT: Goldberg, Robert B.
APPLICANT: Fischer, Robert L.
TITLE OF INVENTION: LEAFY COTYLEDON1 Genes and Their Uses
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

APPLICANT: Harada, John
APPLICANT: Lotan, Tamar
APPLICANT: Ohto, Masa-aki
APPLICANT: Goldberg, Robert B.
APPLICANT: Fischer, Robert L.
TITLE OF INVENTION: LEAFY COTYLEDON1 Genes and Their Uses
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/103,478
FILING DATE: 24-JUN-1998
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 09/026,221
FILING DATE: 19-FEB-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/804,534
FILING DATE: 21-FEB-1997
ATTORNEY/AGENT INFORMATION:
NAME: Einhorn, Gregory P.
REGISTRATION NUMBER: 38,440
REFERENCE/DOCKET NUMBER: 023070-077611US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 27:
SEQUENCE CHARACTERISTICS:
LENGTH: 85 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-103-478-27

Query Match 17.8%; Score 269; DB 4; Length 85;
Best Local Similarity 60.0%; Pred. No. 4.6e-18;
Matches 51; Conservative 18; Mismatches 16; Indels 0; Gaps 0;

QY 41 LPIANVIRMRVLPAAHAKISDDAKETIQCVSEYISFITGEANERQCQRKRTI 100
Db 1 LLPIANVARIMKSPALPENAKISREKQVQCVSEFISFVTGEASEQCTQEKRRKTITGD 60
QY 101 VLWAMSRIGFDDYVEPLGAYLHYR 125
Db 61 VLLALNTLGFENYAEVLKISLTKYR 85

RESULT 12
US-09-103-478-29
Sequence 29, Application US/09103478
Patent No. 6235975
GENERAL INFORMATION:
APPLICANT: Harada, John
APPLICANT: Lotan, Tamar
APPLICANT: Ohto, Masa-aki
APPLICANT: Goldberg, Robert B.
APPLICANT: Fischer, Robert L.
TITLE OF INVENTION: LEAFY COTYLEDON1 Genes and Their Uses
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP

; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/103,478
 ; FILING DATE: 24-JUN-1998
 ; CLASSIFICATION: 800
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 09/026,221
 ; FILING DATE: 19-FEB-1998
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/804,534
 ; FILING DATE: 21-FEB-1997
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Einhorn, Gregory P.
 ; REGISTRATION NUMBER: 38,440
 ; REFERENCE/DOCKET NUMBER: 023070-077611US
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (415) 576-0200
 ; TELEFAX: (415) 576-0300
 ; INFORMATION FOR SEQ ID NO: 23:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 57 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS:
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: peptide
 US-09-103-478-23

Query Match 13.7%; Score 208; DB 4; Length 57;
 Best Local Similarity 66.7%; Pred. No. 1.2e-12;
 Matches 38; Conservative 9; Mismatches 10; Indels 0; Gaps 0;
 QY 69 IOECVSEVIFITGEANERCORQRTITAEVDLWMSRLGFDYVEPLGAYLHRYR 125
 Db 1 VQECVSEFISFITSEASERCHQKRTKINGEDILFAMSTLGFQSYVEPLKLYLQKER 57

RESULT 14
 US-08-681-812-7
 ; Sequence 7, Application US/08681812
 ; Patent No. 5763593
 ; GENERAL INFORMATION:
 ; APPLICANT: Young, Richard A.
 ; APPLICANT: Gadbois, Ellen L.
 ; APPLICANT: Chao, David M.
 ; TITLE OF INVENTION: TBP-Associated Global Negative Regulator
 ; NUMBER OF SEQUENCES: 7
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
 ; STREET: Two Militia Drive
 ; CITY: Lexington
 ; STATE: Massachusetts
 ; COUNTRY: USA
 ; ZIP: 02173
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/681,812
 ; FILING DATE: 29-JUL-1996
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Granahan, Patricia
 ; REGISTRATION NUMBER: 32,227
 ; REFERENCE/DOCKET NUMBER: WHI96-07
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 617-861-6240
 ; TELEFAX: 617-861-9540
 ; INFORMATION FOR SEQ ID NO: 7:
 ; SEQUENCE CHARACTERISTICS:

; LENGTH: 156 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 US-08-681-812-7

Query Match 10.6%; Score 160; DB 1; Length 156;
 Best Local Similarity 31.5%; Pred. No. 1.2e-07;
 Matches 34; Conservative 28; Mismatches 38; Indels 8; Gaps 2;
 QY 37 EQDRLMPIANVIRMRVLPAAKISDDAKETIOECVSEVIFITGEANERCORQRTI 96
 Db 4 DDDLTIPRAAINMKIKETLP-NVRVANDARELVVNCCTEFILHLSSEANEICNKSEKRTI 62
 QY 97 TAEVDLWMSRLGFDYVEPLGAYLHRYREFEGDARGVGLVFGCAAFSR 144
 Db 63 SPEHVIALESGLF-----GSYISEVKEVLQECKTVALKRRKASSR 103

RESULT 15
 US-09-041-886-23
 ; Sequence 23, Application US/09041886
 ; Patent No. 6235872
 ; GENERAL INFORMATION:
 ; APPLICANT: Bredesen, Dale E.
 ; APPLICANT: Rabizadeh, Sharoz
 ; TITLE OF INVENTION: Proapoptotic Peptides, Dependence
 ; NUMBER OF SEQUENCES: 72
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Campbell & Flores LLP
 ; STREET: 4370 La Jolla Village Drive, Suite 700
 ; CITY: San Diego
 ; STATE: California
 ; COUNTRY: United States
 ; ZIP: 92122
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/041,886
 ; FILING DATE:
 ; CLASSIFICATION:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Campbell, Cathryn A.
 ; REGISTRATION NUMBER: 31,815
 ; REFERENCE/DOCKET NUMBER: P-LJ 2626
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (619) 535-9001
 ; TELEFAX: (619) 535-8949
 ; INFORMATION FOR SEQ ID NO: 23:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1185 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 US-09-041-886-23

Query Match 9.7%; Score 147; DB 4; Length 1185;
 Best Local Similarity 31.7%; Pred. No. 2.7e-05;
 Matches 45; Conservative 11; Mismatches 56; Indels 30; Gaps 7;
 QY 149 HPHMSPA--AMLKSRGPFVSGAAMLPHHHHHDMQMHAAAM-----YGGTAVPPPPGPPH 200
 Db 458 HPGFPFPPSTGAQSTAHPPVS-----THHHHHQQQQQQQQQQQQHHGNSGPPPPGAPPH 512
 QY 201 HGGFLMPHPQSSHYL-PYAEPTYGGEHMAAYYGAAYAPNGNGSGDGSGGGG--- 256

Fri Oct 26 13:40:50 2001

Db 513 -----PLEGGSSHHAPYAMSPSLG---SLRPYPGPAHLPPPHSQVYSQAGNGPPV 563
Qy 257 -----GSASHTPOGSGGLEHHP 274
Db 564 SSSSSSSSTSGSYPCSHPS 585

Search completed: October 25, 2001, 09:21:43
Job time: 133 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 25, 2001, 09:16:40 ; Search time 20.67 Seconds
(without alignments)
815.359 Million cell updates/sec

Title: US-09-435-054-2

Perfect score: 1514

Sequence: 1 MDSSFLPAAGNAGSAGG.....ASHTPQSGGLEHPPHPPAYK 278

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_0601.*
1: /SID58/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SID58/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SID58/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SID58/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SID58/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SID58/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SID58/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SID58/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SID58/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SID58/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SID58/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SID58/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SID58/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SID58/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SID58/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SID58/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SID58/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SID58/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SID58/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SID58/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SID58/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SID58/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1514	100.0	278	21	Maize LEC1 # 1. Z
2	516.5	34.1	355	21	Soybean LEC1 # 2 p
3	516.5	34.1	373	21	Soybean LEC1 # 2 p
4	508	33.6	240	21	Soybean LEC1 # 1.
5	494	32.6	208	19	Arabidopsis leafy-
6	494	32.6	208	21	Arabidopsis leafy-
7	489.5	32.3	280	21	Amino acid sequenc
8	484	32.0	214	21	Wheat LEC1 Tril
9	444	29.3	146	21	Veronia mespilifol
10	443	29.3	171	21	Argemone mexicana
11	420.5	27.8	219	21	Soybean LEC1 # 3.
					Pinus radiata tran

12 409 27 0 21 AAY96225 Arabidopsis LEC1 c
13 376.5 24.9 141 21 AAG04651 Arabidopsis thalia
14 360.5 23.8 138 21 AAG31928 Arabidopsis thalia
15 360.5 23.8 139 21 AAG49286 Arabidopsis thalia
16 324.5 21.4 207 12 AAR15481 NP-YB. Homo sapie
17 324.5 21.4 210 21 AAR15481 Human prostate can
18 309.5 20.4 206 21 AAY96219 Arabidopsis thalia
19 303 20.0 107 21 AAY96219 Arabidopsis thalia
20 303 20.0 108 21 AAY96219 Arabidopsis thalia
21 297 19.6 125 21 AAB32808 Pinus radiata tran
22 293.5 19.4 108 21 AAB33041 Eucalyptus grandis
23 287 19.0 104 21 AAG31929 Pinus radiata tran
24 287 19.0 105 21 AAG49287 Arabidopsis thalia
25 235 15.5 65 21 AAY96223 Arabidopsis thalia
26 226 14.9 94 21 AAB32851 Plant LEC1 consens
27 200 13.2 88 21 AAB33084 Eucalyptus grandis
28 199 13.1 78 21 AAB32697 Pinus radiata tran
29 197 13.0 160 21 AAG08107 Arabidopsis thalia
30 192 12.7 163 21 AAG08106 Arabidopsis thalia
31 192 12.7 160 21 AAY96218 Maize LEC1 # 2. Z
32 185.5 12.3 155 21 AAG45969 Arabidopsis thalia
33 185.5 12.3 158 21 AAG45968 Arabidopsis thalia
34 182.5 12.1 400 20 AAY31232 Human c-Maf protei
35 181 12.0 143 21 AAG08108 Arabidopsis thalia
36 173 11.4 117 21 AAB33004 Pinus radiata tran
37 171 11.3 93 21 AAG12405 Zea mays protein f
38 171 11.3 96 21 AAG12404 Zea mays protein f
39 171 11.3 98 20 AAY00942 Soybean Drl protei
40 171 11.3 153 21 AAG12403 Zea mays protein f
41 169.5 11.2 138 21 AAG45970 Arabidopsis thalia
42 165.5 10.9 186 22 AAB63243 Human breast cance
43 153 10.1 113 20 AAY00943 Wheat Drl protein
44 147 9.7 1185 20 AAY33497 Human atrophin I p
45 144 9.5 93 21 AAG01857 Human secreted pro

ALIGNMENTS

RESULT 1
AAY96214
ID AAY96214 standard; Protein: 278 AA.
XX
AC AAY96214;
XX
DT 11-SEP-2000 (first entry)
XX
DE Maize LEC1 # 1.
XX
KW Maize; leafy cotyledon 1 transcriptional activator; LEC1; apomixis;
KW selectable marker; transgenic plant; transgenic seed; HAP3.
XX
OS Zea mays.
XX
PN WO200028058-A2.
XX
PD 18-MAY-2000.
XX
PF 09-NOV-1999; 99WO-US26514.
XX
PR 09-NOV-1998; 98US-0107643.
PR 10-NOV-1998; 98US-0107810.
XX
PA (PION-) PIONEER HI-BRED INT INC.
XX (DUPO-) DU PONT DE NEMOURS & CO E I.
XX
PI Lowe KS, Gordon-Kamm WJ, Klein TM, Rasco-Gaunt S, Cahoon RE;
PI Sun X, Hoerster GJ, Gregory CA, Nadimpalli R;
XX WPI; 2000-376568/32.
DR N-PSDB; AAA27450.
XX
PT New HAP3-type CCAAT-box binding transcriptional activators,

PT particularly leafy cotyledon 1 transcriptional activator, useful for
 XX inducing somatic embryogenesis or apomixis in a plant cell -
 XX
 PS Claim 13; Page 77-78; 94pp; English.
 XX
 CC The present sequence is the maize leafy cotyledon 1 transcriptional
 CC activator, LEC1. This sequence is a HAP3-type CCAAT-box binding protein.
 CC LEC1 expression initiates the formation of embryo-like structures and
 CC improves growth and recovery of transformants. When apomixis occurs, i.e.
 CC the replacement of sexual reproduction by asexual reproduction, LEC1
 CC expression in the nucellus integument, or cell specific expression in
 CC the megaspore mother cell would trigger embryo formation from maternal
 CC tissues only. This results in the production of seeds identical to the
 CC parent. Using LEC1, transgenic high yielding seeds could be developed. In
 CC addition, LEC1 could be used for positive selection of a transformed cell
 CC (transgenic plant), for increasing transformation efficiency and for
 CC increasing recovery of regenerated plants.
 XX
 SQ Sequence 278 AA;

Query Match 100.0%; Score 1514; DB 21; Length 278;
 Best Local Similarity 100.0%; Pred. No. 3.3e-127;
 Matches 278; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDSSSFLPAAGAEAGSAGGANNNGAAQQAAPAIHQDRLMPIANVIRMRVLPAAHAK 60
 DB 1 MDSSSFLPAAGAEAGSAGGANNNGAAQQAAPAIHQDRLMPIANVIRMRVLPAAHAK 60
 QY 61 ISDDAKETIQECVSEYISFITGEANERCQRTTAEVDLWMSRLGDFDYVEPLGAY 120
 DB 61 ISDDAKETIQECVSEYISFITGEANERCQRTTAEVDLWMSRLGDFDYVEPLGAY 120
 QY 121 LHYRFEFGDARGVLPGAAPSRGDDHPSMAAMLSKRGVSGAAMLPHHHHHDM 180
 DB 121 LHYRFEFGDARGVLPGAAPSRGDDHPSMAAMLSKRGVSGAAMLPHHHHHDM 180
 QY 181 QMHAAMYGGTAVPPAGPHGGFLMPHFGSSHYLPYAEPTYGGEHAMAAYYGAAYA 240
 DB 181 QMHAAMYGGTAVPPAGPHGGFLMPHFGSSHYLPYAEPTYGGEHAMAAYYGAAYA 240
 QY 241 PNGSGDGGSGGGGSGASHTPQSGGLEHPPHFAK 278
 DB 241 PNGSGDGGSGGGGSGASHTPQSGGLEHPPHFAK 278

RESULT 2
 AAY96224
 ID AAY96224 standard; Protein; 355 AA.
 XX
 AC AAY96224;

11-SEP-2000 (first entry)
 XX
 DE Soybean LEC1 # 2 protein # 2.
 XX
 KW Soybean; leafy cotyledon 1 transcriptional activator; LEC1; apomixis;
 KW selectable marker; transgenic plant; transgenic seed; HAP3.
 XX
 OS Glycine max.
 XX
 PN WO200028058-A2.
 XX
 PD 18-MAY-2000.
 XX
 PF 09-NOV-1999; 99WO-US26514.
 XX
 PR 09-NOV-1998; 98US-0107643.
 PR 10-NOV-1998; 98US-0107810.
 XX
 XX (PION-) PIONEER HI-BRED INT INC.
 PA (DUPO) DU PONT DE NEMOURS & CO E. I.
 XX

PI Lowe KS, Gordon-Kamm WJ, Klein TM, Rasco-Gaunt S, Cahoon RE;
 XX Sun X, Hoerster GJ, Gregory CA, Nadimpalli R;
 XX WPI: 2000-376568/32.
 DR N-PSDB; AAA27460.
 XX
 PT New HAP3-type CCAAT-box binding transcriptional activators,
 PT particularly leafy cotyledon 1 transcriptional activator, useful for
 PT inducing somatic embryogenesis or apomixis in a plant cell -
 XX
 XX Claim 13; Page 89-90; 94pp; English.
 XX
 CC The present sequence is the soybean leafy cotyledon 1 transcriptional
 CC activator, LEC1. This sequence is a HAP3-type CCAAT-box binding protein.
 CC LEC1 expression initiates the formation of embryo-like structures and
 CC improves growth and recovery of transformants. When apomixis occurs,
 CC i.e. the replacement of sexual reproduction by asexual reproduction,
 CC LEC1 expression in the nucellus integument, or cell specific expression
 CC in the megaspore mother cell would trigger embryo formation from
 CC maternal tissues only. This results in the production of seeds identical
 CC to the parent. Using LEC1, transgenic high yielding seeds could be
 CC developed. In addition, LEC1 could be used for positive selection of a
 CC transformed cell (transgenic plant), for increasing transformation
 CC efficiency and for increasing recovery of regenerated plants.
 CC Note: this sequence is different from that of AAY96220, even though
 CC they are both encoded by the same nucleotide sequence (AAA27460). The
 CC AAY96220 sequence has 18 stop codons distributed throughout the CDS,
 CC while the protein of AAY96224 has the stop codons omitted.
 XX
 SQ Sequence 355 AA;

Query Match 34.1%; Score 516.5; DB 21; Length 355;
 Best Local Similarity 53.8%; Pred. No. 3.3e-38;
 Matches 113; Conservative 16; Mismatches 50; Indels 31; Gaps 4;

QY 16 SNAGGANNNGAAQQAAPAIHQDRLMPIANVIRMRVLPAAHAKISDDAKETIQECVSE 75
 DB 38 ssdqncshsaageenectvreqdfrfmbianvirimrkilpphakisdaketi qecvse 97
 QY 76 YISFITGEANERCQRTTAEVDLWMSRLGDFDYVEPLGAYLHRYRFEFGDARGV 135
 DB 98 ylsfitgeanercqqrktitaedvlwamsklgfdyiepltmlylhryrelegdrtsm- 156
 QY 136 LVPGAAPSRGDDHPSMAAMLSKRGVSGAAMLPHHHHHDMQMAAMYGGTAV--- 192
 DB 157 -----rgeplgkrtveyatl-----atafppppfhhhngyfgaampmgtyvret 200
 QY 193 PPPAGPHGGFLMPHFGSSHYLPYAYEP 222
 DB 201 ppnaasshh-----hhgisnahep 219

RESULT 3
 AAY96220
 ID AAY96220 standard; Protein; 373 AA.
 XX
 AC AAY96220;
 XX
 DT 11-SEP-2000 (first entry)
 XX
 DE Soybean LEC1 # 2 protein # 1.
 XX
 KW Soybean; leafy cotyledon 1 transcriptional activator; LEC1; apomixis;
 KW selectable marker; transgenic plant; transgenic seed; HAP3.
 XX
 OS Glycine max.
 XX
 XX Key Location/Qualifiers
 FT Misc-difference 1..373
 FT /note= "All Xs are unspecified residues encoded by
 FT stop codons"
 XX

FN WO200028058-A2.
 XX 18-MAY-2000.
 PD 09-NOV-1999; 99WO-US26514.
 XX 09-NOV-1998; 98US-0107643.
 XX 10-NOV-1998; 98US-0107810.
 XX (PION-) PIONEER HI-BRED INT INC.
 PA (DUPO) DU PONT DE NEMOURS & CO E I.
 XX Lowe KS, Gordon-Kamm WJ, Klein TM, Rasco-Gaunt S, Cahoon RE;
 PI Sun X, Hoerster GJ, Gregory CA, Nadimpalli R;
 DR N-PSDB; AAA27460.
 XX New HAP3-type CCAAT-box binding transcriptional activators,
 XX particularly leafy cotyledon 1 transcriptional activator, useful for
 XX inducing somatic embryogenesis or apomixis in a plant cell -
 PS Claim 13; Page 89-90; 94pp; English.
 XX The present sequence is the soybean leafy cotyledon 1 transcriptional
 CC activator, LEC1. This sequence is a HAP3-type CCAAT-box binding protein.
 CC LEC1 expression initiates the formation of embryo-like structures and
 CC improves growth and recovery of transformants. When apomixis occurs,
 CC i.e. the replacement of sexual reproduction by asexual reproduction,
 CC LEC1 expression in the nucleellus integument, or cell specific expression
 CC in the megaspore mother cell would trigger embryo formation from
 CC maternal tissues only. This results in the production of seeds identical
 CC to the parent. Using LEC1, transgenic high yielding seeds could be
 CC developed. In addition, LEC1 could be used for positive selection of a
 CC transformed cell (transgenic plant), for increasing transformation
 CC efficiency and for increasing recovery of regenerated plants.
 CC Note: this sequence is different from that of AAY96224, even though
 CC they are both encoded by the same nucleotide sequence (AAZ7460). The
 CC AAY96220 sequence has 18 stop codons distributed throughout the CDS,
 CC while the protein of AAY96224 has the stop codons omitted.
 XX Sequence 373 AA;
 SQ
 Query Match 34.1%; Score 516.5; DB 21; Length 373;
 Best Local Similarity 53.8%; Pred. No. 3.5e-38;
 Matches 113; Conservative 16; Mismatches 50; Indels 31; Gaps 4;
 QY 16 SAAGGANGGAQAQHPAIREODRLMPTIANVIRIMRRLVPAHAKISDDAKETIQECVSE 75
 Db 38 ssdncnshsaageenectvreqdrfmpianvirimrkilpphakisddaketi qcvse 97
 QY 76 YISFITGEANERCQQRKTTAEDVLWAMSRGLGFDYVEPLGAYLHRYREPEGDARGVG 135
 Db 98 yisfitgeanercqrkittaedvlwamskigfdyieplumylhryrelegdrtsm- 156
 QY 136 LVPGAAPSRGDDHHPHSMSPAAMLKSRGPVSGAAMLPHHHHHHDMQMHAAWYGGTAV--- 192
 Db 157 -----rgeplgkrtyeatl-----atafvppfhbhngyfgaampmgtyvret 200
 QY 193 PPPAGPPHGGFLMPHPQGSSHYLPYAYEP 222
 Db 201 ppaasshh-----hghisnahep 219
 RESULT 4
 AAY96216
 ID AAY96216 standard; Protein; 240 AA.
 XX AAY96216;
 XX
 XX 11-SEP-2000 (first entry)
 XX

DE Soybean LEC1 # 1.
 XX Soybean; leafy cotyledon 1 transcriptional activator; LEC1; apomixis;
 KW selectable marker; transgenic plant; transgenic seed; HAP3.
 XX Glycine max.
 OS WO200028058-A2.
 XX 18-MAY-2000.
 PD 09-NOV-1999; 99WO-US26514.
 XX 09-NOV-1998; 98US-0107643.
 PR 10-NOV-1998; 98US-0107810.
 XX (PION-) PIONEER HI-BRED INT INC.
 PA (DUPO) DU PONT DE NEMOURS & CO E I.
 XX Lowe KS, Gordon-Kamm WJ, Klein TM, Rasco-Gaunt S, Cahoon RE;
 PI Sun X, Hoerster GJ, Gregory CA, Nadimpalli R;
 DR WPI; 2000-376568/32.
 DR N-PSDB; AAA27456.
 XX New HAP3-type CCAAT-box binding transcriptional activators,
 XX particularly leafy cotyledon 1 transcriptional activator, useful for
 XX inducing somatic embryogenesis or apomixis in a plant cell -
 PS Claim 13; Page 82; 94pp; English.
 XX The present sequence is the soybean leafy cotyledon 1 transcriptional
 CC activator, LEC1. This sequence is a HAP3-type CCAAT-box binding protein.
 CC LEC1 expression initiates the formation of embryo-like structures and
 CC improves growth and recovery of transformants. When apomixis occurs, i.e.
 CC the replacement of sexual reproduction by asexual reproduction, LEC1
 CC expression in the nucleellus integument, or cell specific expression in
 CC the megaspore mother cell would trigger embryo formation from maternal
 CC tissues only. This results in the production of seeds identical to the
 CC parent. Using LEC1, transgenic high yielding seeds could be developed. In
 CC addition, LEC1 could be used for positive selection of a transformed cell
 CC (transgenic plant), for increasing transformation efficiency and for
 CC increasing recovery of regenerated plants.
 XX Sequence 240 AA;
 SQ
 Query Match 33.6%; Score 508; DB 21; Length 240;
 Best Local Similarity 52.5%; Pred. No. 1.2e-37;
 Matches 114; Conservative 19; Mismatches 52; Indels 32; Gaps 5;
 QY 10 AGAENGSAAGGANGGAQAQHPAIREODRLMPTIANVIRIMRRLVPAHAKISDDAKETI 69
 Db 47 asdsaaatgeene-----ctvreqdrfmpianvirimrkilpphakisddaketi 97
 QY 70 QECVSEYISFITGEANERCQQRKTTAEDVLWAMSRGLGFDYVEPLGAYLHRYREPEG 129
 Db 98 qecvseyisfitgeanercqrkittaedvlwamskigfdyieplumylhryreleg 157
 QY 130 DARGVGLVPGAAPSRGDDHHPHSMSPAAMLKSRGPVSGAAMLPHHHHHHDMQMHAAWYGG 189
 Db 158 drtsm-----rgeplgkrtyeatl-----gvatafvpppyhhbhngyfgaampmg 202
 QY 190 T-----AVPPAGPPHGGFLMPHPQGSSHYLPYAYEP 222
 Db 203 tyvreappntasshhhhhhhhhhhhargisn-----ahap 235
 RESULT 5
 AAW71722
 ID AAW71722 standard; Protein; 208 AA.
 XX
 XX AAW71722;
 AC


```

Db 68 yisvtgeanercqrekrkitaediilwamskigfdnyvqpltvfnryreietd-rgsa 126
QY 136 LVPGAAPSRGDHPHSMSPAAMLKSRGPVSGAAMLPHHHHHHDMQMHAAWYGTAVPPP 195
Db 127 l-----rg-----eppsrlqt-----ygnngl--- 143
QY 196 AGPHHGGFLMPHQSSHYL-----PYAYEPTYGGEHMAAYGGAAYAFNGSGDGGSG 251
Db 144 -----gf-----hpshtglpppgpyg-----gmldqsmvnggryyq-ngsssgdgs 185
QY 252 SGGGGGSAS 260
Db 186 svggsgssss 194

RESULT 7
ID AAY96222 standard; Protein; 280 AA.
AC AAY96222;
XX
XX 11-SEP-2000 (first entry)
XX Wheat LEC1.
XX
XX DE
XX KW wheat; leafy cotyledon 1 transcriptional activator; LEC1; apomixis;
XX KW selectable marker; transgenic plant; transgenic seed; HAP3.
XX OS
XX PN Triticum aestivum.
XX PD WO200028058-A2.
XX 18-MAY-2000.
XX 09-NOV-1999; 99WO-US26514.
XX 09-NOV-1998; 98US-0107643.
XX 10-NOV-1998; 98US-0107810.
XX (PION-) PIONEER HI-BRED INT INC.
XX (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
XX PI Lowe KS, Gordon-Kamm WJ, Klein TM, Rasco-Gaunt S, Cahoon RE;
XX PI Sun X, Hoerster GJ, Gregory CA, Nadimpalli R;
XX
XX DR WPI: 2000-376568/32.
XX DR N-PSDB; AAA27462.
XX
XX PT New HAP3-type CCAAT-box binding transcriptional activators,
XX PT particularly leafy cotyledon 1 transcriptional activator, useful for
XX PT inducing somatic embryogenesis or apomixis in a plant cell -
XX
XX PS Claim 13; Page 93; 94pp; English.
XX
XX CC The present sequence is the wheat leafy cotyledon 1 transcriptional
XX CC activator, LEC1. This sequence is a HAP3-type CCAAT-box binding protein.
XX CC LEC1 expression initiates the formation of embryo-like structures and
XX CC improves growth and recovery of transformation by asexual reproduction, i.e.
XX CC the replacement of sexual reproduction by asexual reproduction, LEC1
XX CC expression in the nucellus integument, or cell specific expression in
XX CC the megaspore mother cell would trigger embryo formation from maternal
XX CC tissues only. This results in the production of seeds identical to the
XX CC parent. Using LEC1, transgenic high yielding seeds could be developed. In
XX CC addition, LEC1 could be used for positive selection of a transformed cell
XX CC (transgenic plant), for increasing transformation efficiency and for
XX CC increasing recovery of regenerated plants.
XX
XX SQ Sequence 280 AA;

```

Query Match 32.38; Score 489.5; DB 21; Length 280;
 Best Local Similarity 41.5%; Pred. No. 6.3e-36;

```

Matches 119; Conservative 27; Mismatches 92; Indels 49; Gaps 9;
QY 11 GAENSGAAGGANNNGAQAQAAPAIREDRLMPIANVIRIMRVLPAAKISDDAKETIQ 70
Db 5 gypngpaapaptqg-----tpvvreqdrmpianvirimralpahakisddakeaiq 57
QY 71 ECVSEVISFITGEANERCOEORKTITAEEDVLWMSRLGFDYVEPLGAYLHRYREFE-- 128
Db 58 ecvsefistvgeanercrmqhrktvnaedivwalnrgfdyvvplsvflhrmdpeag 117
QY 129 -----GDARGVCLVPGAAPSRGSDHHPHSM-----SPAAMLKSRGPVSGAAMLPH 173
Db 118 tggaaagdsrav-----tsappraappviahvplqagrpmyappaplvqenmqtrpyap- 172
QY 174 HHHHHDMQMHAAWYGTAVPPPAGPPHNG---GFLMPHQSSSHYLYAYEPTYGGEHAM 230
Db 173 papvqvqmqrglygpra-----pvhgyavgmavpranvnggy-----qvfgegv 218
QY 231 AAYGGAAYAFNGSGDGGSGGGG---GSASHTPPQSGGLEHHPHP 274
Db 219 aqyygygyeagaygagssnggaalgdeessngvvpagpgmgep 265

RESULT 8
AAY96217
ID AAY96217 standard; Protein; 214 AA.
XX
XX AC AAY96217;
XX
XX DT 11-SEP-2000 (first entry)
XX
XX DE Veronia mespilifolia LEC1.
XX KW Leafy cotyledon 1 transcriptional activator; LEC1; apomixis;
XX KW selectable marker; transgenic plant; transgenic seed; HAP3.
XX OS Veronia mespilifolia.
XX PN WO200028058-A2.
XX PD 18-MAY-2000.
XX
XX PF 09-NOV-1999; 99WO-US26514.
XX
XX PR 09-NOV-1998; 98US-0107643.
XX PR 10-NOV-1998; 98US-0107810.
XX
XX PA (PION-) PIONEER HI-BRED INT INC.
XX PA (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
XX PI Lowe KS, Gordon-Kamm WJ, Klein TM, Rasco-Gaunt S, Cahoon RE;
XX PI Sun X, Hoerster GJ, Gregory CA, Nadimpalli R;
XX
XX DR WPI: 2000-376568/32.
XX DR N-PSDB; AAA27457.
XX
XX PT New HAP3-type CCAAT-box binding transcriptional activators,
XX PT particularly leafy cotyledon 1 transcriptional activator, useful for
XX PT inducing somatic embryogenesis or apomixis in a plant cell -
XX
XX PS Claim 13; Page 83-84; 94pp; English.
XX
XX CC The present sequence is the leafy cotyledon 1 transcriptional activator,
XX CC LEC1 of Veronia mespilifolia. This sequence is a HAP3-type CCAAT-box
XX CC binding protein. LEC1 expression initiates the formation of embryo-like
XX CC structures and improves growth and recovery of transformation. During
XX CC apomixis, i.e. the replacement of sexual reproduction by asexual
XX CC reproduction, LEC1 expression in the nucellus integument, or cell
XX CC specific expression in the megaspore mother cell triggers embryo
XX CC formation from maternal tissues only. This results in the production of
XX CC seeds identical to the parent. Using LEC1, transgenic high yielding seeds
XX CC could be developed. In addition, LEC1 could be used for positive
XX CC selection of a transformed cell (transgenic plant), for increasing

```


XX Example 6; Fig 1; 94pp; English.
 PS The present sequence is the leafy cotyledon 1 transcriptional activator,
 CC LEC1 consensus sequence from Arabidopsis. This sequence is a HAP3-type
 CC CCAAT-box binding protein. LEC1 expression initiates the formation of
 CC embryo-like structures and improves growth and recovery of
 CC transformants. When apomixis occurs, i.e. the replacement of sexual
 CC reproduction by asexual reproduction, LEC1 expression in the nucellus
 CC integument, or cell specific expression in the megaspore mother cell
 CC would trigger embryo formation from maternal tissues only. This results
 CC in the production of seeds identical to the parent. Using LEC1,
 CC transgenic high yielding seeds could be developed. In addition, LEC1
 CC could be used for positive selection of a transformed cell (transgenic
 CC plant), for increasing transformation efficiency and for increasing
 CC recovery of regenerated plants. The present sequence was used in the
 CC identification of the plant LEC1 consensus sequence (AAY96223).
 XX
 SQ Sequence 90 AA;

Query Match 27.0%; Score 409; DB 21; Length 90;
 Best Local Similarity 83.3%; Pred. No. 2.4e-29;
 Matches 75; Conservative 11; Mismatches 4; Indels 0; Gaps 0;
 QY 36 REODRLPIANVIRIMRRVLPAAKISDDAKETIQECVSEYISFTIGEANRCQREQRT 95
 Db 1 reqdgmplanvirimrktlpshakisdaketi qecvseyisftigeanrcqreqrt 60
 QY 96 ITAEDVLWMSRLGFDYDVEPLGAYLHYR 125
 Db 61 itaedilwamsklgfdydvplvfinryr 90

RESULT 13
 AAG04651
 ID AAG04651 standard; Protein; 141 AA.
 XX
 AC AAG04651;
 XX
 DT 17-OCT-2000 (first entry)
 XX
 DE Arabidopsis thaliana protein fragment SEQ ID NO: 757.
 KW Protein identification; signal transduction pathway; metabolic pathway;
 KW hybridisation assay; genetic mapping; gene expression control; promoter;
 KW termination sequence.
 XX
 OS Arabidopsis thaliana.
 XX
 PN EPI033405-A2.
 XX
 PD 06-SEP-2000.
 XX
 PF 25-FEB-2000; 2000EP-0301439.
 XX
 PR 25-FEB-1999; 99US-0121825.
 PR 05-MAR-1999; 99US-0123180.
 PR 09-MAR-1999; 99US-0123548.
 PR 23-MAR-1999; 99US-0125788.
 PR 25-MAR-1999; 99US-0126264.
 PR 29-MAR-1999; 99US-0126785.
 PR 01-APR-1999; 99US-0127462.
 PR 06-APR-1999; 99US-0128234.
 PR 08-APR-1999; 99US-0128714.
 PR 16-APR-1999; 99US-0129845.
 PR 19-APR-1999; 99US-0130077.
 PR 21-APR-1999; 99US-0130449.
 PR 23-APR-1999; 99US-0130510.
 PR 28-APR-1999; 99US-0130891.
 PR 30-APR-1999; 99US-0131449.
 PR 30-APR-1999; 99US-0132048.
 PR 30-APR-1999; 99US-0132407.

PR 04-MAY-1999; 99US-0132484.
 PR 05-MAY-1999; 99US-0132485.
 PR 06-MAY-1999; 99US-0132486.
 PR 07-MAY-1999; 99US-0132487.
 PR 11-MAY-1999; 99US-0132863.
 PR 14-MAY-1999; 99US-0134256.
 PR 14-MAY-1999; 99US-0134218.
 PR 14-MAY-1999; 99US-0134219.
 PR 14-MAY-1999; 99US-0134221.
 PR 14-MAY-1999; 99US-0134370.
 PR 18-MAY-1999; 99US-0134768.
 PR 19-MAY-1999; 99US-0134941.
 PR 20-MAY-1999; 99US-0135124.
 PR 21-MAY-1999; 99US-0135353.
 PR 24-MAY-1999; 99US-0135629.
 PR 25-MAY-1999; 99US-0136021.
 PR 27-MAY-1999; 99US-0136392.
 PR 28-MAY-1999; 99US-0136782.
 PR 01-JUN-1999; 99US-0137222.
 PR 03-JUN-1999; 99US-0137528.
 PR 04-JUN-1999; 99US-0137502.
 PR 07-JUN-1999; 99US-0137724.
 PR 08-JUN-1999; 99US-0138094.
 PR 10-JUN-1999; 99US-0138540.
 PR 10-JUN-1999; 99US-0138847.
 PR 14-JUN-1999; 99US-0139119.
 PR 16-JUN-1999; 99US-0139452.
 PR 16-JUN-1999; 99US-0139453.
 PR 17-JUN-1999; 99US-0139492.
 PR 18-JUN-1999; 99US-0139454.
 PR 18-JUN-1999; 99US-0139455.
 PR 18-JUN-1999; 99US-0139456.
 PR 18-JUN-1999; 99US-0139457.
 PR 18-JUN-1999; 99US-0139458.
 PR 18-JUN-1999; 99US-0139459.
 PR 18-JUN-1999; 99US-0139460.
 PR 18-JUN-1999; 99US-0139461.
 PR 18-JUN-1999; 99US-0139462.
 PR 18-JUN-1999; 99US-0139463.
 PR 18-JUN-1999; 99US-0139750.
 PR 18-JUN-1999; 99US-0139763.
 PR 21-JUN-1999; 99US-0139817.
 PR 22-JUN-1999; 99US-0139899.
 PR 23-JUN-1999; 99US-0140353.
 PR 23-JUN-1999; 99US-0140354.
 PR 24-JUN-1999; 99US-0140695.
 PR 28-JUN-1999; 99US-0140823.
 PR 29-JUN-1999; 99US-0140991.
 PR 30-JUN-1999; 99US-0141287.
 PR 01-JUL-1999; 99US-0141842.
 PR 01-JUL-1999; 99US-0142154.
 PR 02-JUL-1999; 99US-0142055.
 PR 06-JUL-1999; 99US-0142390.
 PR 08-JUL-1999; 99US-0142803.
 PR 09-JUL-1999; 99US-0142920.
 PR 12-JUL-1999; 99US-0142977.
 PR 13-JUL-1999; 99US-0143542.
 PR 14-JUL-1999; 99US-0143624.
 PR 15-JUL-1999; 99US-0144005.
 PR 16-JUL-1999; 99US-0144085.
 PR 16-JUL-1999; 99US-0144086.
 PR 19-JUL-1999; 99US-0144325.
 PR 19-JUL-1999; 99US-0144331.
 PR 19-JUL-1999; 99US-0144332.
 PR 19-JUL-1999; 99US-0144333.
 PR 19-JUL-1999; 99US-0144334.
 PR 19-JUL-1999; 99US-0144335.
 PR 20-JUL-1999; 99US-0144352.
 PR 20-JUL-1999; 99US-0144632.
 PR 20-JUL-1999; 99US-0144884.
 PR 21-JUL-1999; 99US-0144814.
 PR 21-JUL-1999; 99US-0145086.
 PR 21-JUL-1999; 99US-0145088.

```
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 22-JUL-1999; 99US-0145194.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149502.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150366.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151130.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.

PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.
PR 29-OCT-1999; 99US-0162142.

Query Match 24.9%; Score 376.5; DB 21; Length 141;
Best Local Similarity 46.0%; Pred. No. 3.2e-26;
Matches 75; Conservative 29; Mismatches 30; Indels 29; Gaps 3;

QY 8 PAAGAENGSAAGGANNNGAAQQAAPATREODRLMPIANVIRIMRRVLPAAHAKISDDAKE 67
Db 5 pspagdggesg-----:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 68 TIQECVSEYISITGEANERCQORKTITARDVLWMSRLQFDDYVEPLGAYLHRYREF 127
Db 52 tvqecvsefifitseasckckekrktvngdldllwamatigfedypleplkiylaryrel 111
QY 128 EGDARGVGLVPGAAPSRGDDHHPHSMSPAAMLKSRGPVSGAAM 170
Db 112 egdnkgsg-----ksgd-----gsnrdagggvsgeem 138

RESULT 14
AAG31928
ID AAG31928 standard; Protein; 138 AA.
XX
AC AAG31928;
XX
DT 17-OCT-2000 (first entry)
DE Arabidopsis thaliana protein fragment SEQ ID NO: 38427.
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
```

PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135124.
PR 24-MAY-1999; 99US-0135353.
PR 25-MAY-1999; 99US-0135629.
PR 27-MAY-1999; 99US-0136021.
PR 28-MAY-1999; 99US-0136392.
PR 01-JUN-1999; 99US-0136782.
PR 03-JUN-1999; 99US-0137222.
PR 04-JUN-1999; 99US-0137528.
PR 07-JUN-1999; 99US-0137502.
PR 08-JUN-1999; 99US-0137724.
PR 10-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139859.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140981.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144332.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.

PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 13-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0158458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.


```
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.

PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 23.8%; Score 360.5; DB 21; Length 139;
Best Local Similarity 45.4%; Pred. No. 8.4e-25;
Matches 74; Conservative 29; Mismatches 29; Indels 31; Gaps 4;

Qy 8 PAAGAENGAAGANNNGAAQAAPAIHQDRLMPTIANVIRIMRRVLPAAHAKISDDAKE 67
Db 5 pspagdggesgg-----svredrylpianisrlnkkalppngkigkdakd 51
Qy 68 TIQECVSEYISFTIGEANERCQREQRTITAEADVLMASRLGDDYVEPLGAYLHRYREF 127
Db 52 tvqecvsefifitseasdkckekrtvngddllwamatlgfedylepkiylaryr-- 109
Qy 128 EGDARGVGLVPGAAPSRGGDHHPHSMSPAAMLKSRGPVSGNAAM 170
Db 110 egdnkgsg-----ksgd-----gsnrddagggvsggeem 136

Search completed: October 25, 2001, 09:21:24
Job time: 284 sec
```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 25, 2001, 09:20:20 ; Search time 16.39 seconds
(without alignments)
1292.039 Million cell updates/sec

Title: US-09-435-054-2
Perfect score: 1514
Sequence: 1 MDSSFLPAAGNCSAAG.....ASHTPQSGGLEPHPPAYK 278
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues
Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	494	32.6	208	2 G86352	protein T26F17.20
2	410.5	27.1	161	2 G71407	transcription fact
3	410	27.1	228	2 T45874	transcription fact
4	391	25.8	179	2 S22820	transcription fact
5	372	24.6	178	2 A84788	probable CCAAT-box
6	360.5	23.8	138	2 E84810	hypothetical prote
7	357.5	23.6	160	2 G84919	probable CCAAT-box
8	354.5	23.4	215	2 F84508	probable CCAAT-box
9	324.5	21.4	205	2 S22817	transcription fact
10	324.5	21.4	207	2 F38245	transcription fact
11	324.5	21.4	207	2 A23692	transcription fact
12	322	21.3	180	2 S24469	transcription fact
13	320	21.1	186	2 JC6080	transcription fact
14	319.5	21.1	209	2 S22818	transcription fact
15	307	20.3	139	2 C86222	transcription fact
16	295.5	19.5	144	1 A28123	hypothetical prote
17	274	18.1	116	2 S42744	transcription fact
18	265	17.5	205	2 S51565	transcription fact
19	232.5	15.4	122	2 S22819	transcription fact
20	195	12.9	159	2 S53582	transcription fact
21	190.5	12.6	162	2 T50504	transcription fact
22	176.5	11.7	370	2 I37555	TATA-binding prote
23	171	11.3	590	2 A26638	c-Maf protein - mo
24	170	11.2	369	1 TVFVAF	homeotic protein D
25	161	10.6	176	2 A43320	transforming prote
26	160	10.6	175	2 JC5365	TATA-binding prote
27	147.5	9.7	963	2 T48707	TBP-binding repres
28	147	9.7	1184	2 G01763	related to regulat
29	147	9.7	1184	2 S50832	atrophin-1 - human

30	142	9.4	161	2 T40194	probable transcript
31	140.5	9.3	420	2 I59234	octamer binding tr
32	138	9.1	420	2 A49642	transcription fact
33	138	9.1	839	2 F75518	hypothetical prote
34	138	9.1	1212	2 T13804	shs protein - fui
35	138	9.1	1585	2 T31611	hypothetical prote
36	135.5	8.9	561	2 T16148	hypothetical prote
37	134	8.9	427	2 A23372	female-specific do
38	134	8.9	549	2 B23372	annexin VII - Afri
39	132.5	8.8	512	2 S70644	male-specific doub
40	129.5	8.6	200	2 S10334	glycine-rich prote
41	129	8.5	440	2 S71795	transcription fact
42	129	8.5	910	2 A34721	androgen receptor
43	129	8.5	911	2 B34721	androgen receptor
44	129	8.5	1166	2 T13958	syngap-bl protein
45	129	8.5	1249	2 T14270	Ras-GTPase activat

ALIGNMENTS

RESULT 1
G86352
protein T26F17.20 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: G86352
R:Theologos, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon
ansen, N.F.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzia
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719
A:Accession: G86352
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-208 <STO>
A:Cross-references: GB:AE005172; NID:96552738; PIDN:AAF16537.1; GSPDB:GN00141
C:Genetics:
A:Gene: T26F17.20
A:Map position: 1

Query Match 32.6%; Score 494; DB 2; Length 208;
Best Local Similarity 47.0%; Pred. No. 3.2e-29;
Matches 117; Conservative 19; Mismatches 47; Indels 66; Gaps 9;

QY	16	SAAGGANNNGAAQHAAPATREQDLMPPIANVIRMRVLPAAKISDDAKETIQECVSE	75
DB	8	AGADKNGIVVQQPCVAREQQYMPPIANVIRMRKTLPSHAKISDDAKETIQECVSE	67
QY	76	YISFITGEANERCQQRKTIATDVLWMSRLGFDYVPLGAYLHRYREFEGDARGVG	135
DB	68	YISFVTGEANERCQQRKTIATDILWMSKLGFDYVPLVFNRYREIETD-RGSA	126
QY	136	LVPGAAFSRGDHPHSHMSPAAMLSKRGVSGAAMLPFHHDHDMQMAAAYGGTAVPPP	195
DB	127	L-----RG-----EPPSLRQT-----YGGNGI---	143
QY	196	AGPPHGGFLMHPHQQSSHYL----PVAYEPTYGGEHMAAAYGGRAYAPGNGGGSG	251
DB	144	-----GF-----HGFSLGLPPPGPIGY-----GMLDQSMVMWGGRYYO-NGSSGODES	185
QY	252	SGGGGGSGAS	260
DB	186	SVGGGGSSS	194

```

RESULT 2
G71407
transcription factor, CCAAT-binding, chain A - Arabidopsis thaliana
N:Alternate names: protein DL3310W
C:Species: Arabidopsis thaliana (mouse-ear cress)
A:Variety: Columbia
C:Date: 03-Aug-1998 #sequence_revision 03-Aug-1998 #text_change 20-Jun-2000
C:Accession: G71407
R:Bevan, M.; Bancroft, I.; Bent, E.; Love, K.; Goodman, H.; Dean, C.; Bergkamp, R.; Dirk
P.; Wedler, H.; Wedler, E.; Wambutt, R.; Weitzenecker, T.; Pohl, T.M.; Terry, N.; Giel
avanagh, T.; Hempel, S.; Kottler, P.; Entian, K.D.; Kieder, M.; Schaeffer, M.; Funk, B.
Nature 391, 485-488, 1998
A:Authors: Mueller-Auer, S.; Silvey, M.; James, R.; Montfort, A.; Pons, A.; Puigdomenech
erhoff, A.; Moores, T.; Jones, J.D.G.; Eneva, T.; Palme, K.; Benes, V.; Rechman, S.; Ans
C.; Chalwatzis, N.
A:Title: Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of Arabidopsis tha
A:Reference number: A71400; MUID:98121113
C:Accession: G71407
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-161 <REV>
A:Cross-references: GB:Z97336; MID:q2244788; PIDN:CAB10233.1; PID:q2244810
C:Genetics:
A:Map position: 4COP9-4G3845
A:Note: DL3310W
C:Superfamily: transcription factor HAP3
C:Keywords: DNA binding; transcription regulation
F:20-109/Domain: DNA binding #status predicted <DNA>

Query Match 27.1%; Score 410.5; DB 2; Length 161;
Best Local Similarity 48.3%; Pred. No. 3e-23;
Matches 87; Conservative 23; Mismatches 35; Indels 35; Gaps 4;

QY 12 AENGSAAGGANNNGAAQQAAPAIHQDRLMPANVIRMRVLPAAHAKISDDAKETIQE 71
DB 2 ADSNDGKHGDGNA-----STEQDRFLPIANVIRMRVLPAAHAKISDDAKETIQE 55
QY 72 CVSEYISFTTGEANERQORRTITAEVDLWMSRLGFDYVPEPLGAYLHRYREFE--- 128
DB 56 CVSEYISFTTGEASDRKQREKRTINGDGLLWMTTLGFDYVPEPLKYLQYREVEGEK 115
QY 129 -----GDARGVGLVPGAAPSRGGDHHPHSMSPAAMLSKRGVPSGAAM---LPHHHHH 178
DB 116 TTTAGRGQDKEGGGGGGAGSGG-----APMYGGMTVTMGHOFSSH 159

RESULT 3
T45874
transcription factor NF-Y, CCAAT-binding-like protein - Arabidopsis thaliana
N:Alternate names: protein F4P12.40
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 04-Feb-2000
C:Accession: T45874
R:Blöcker, H.; Mewes, H.W.; Lemcke, K.; Mayer, K.F.X.; Quétier, F.; Salanoubat, M.
submitted to the Protein Sequence Database, January 2000
A:Reference number: Z23016
A:Accession: T45874
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-228 <BLO>
A:Cross-references: EMBL:AL132966
A:Experimental source: cultivar Columbia; BAC clone F4P12
C:Genetics:
A:Map position: 3
A:Introns: 75/2; 118/3; 120/3; 145/3; 164/3; 179/3; 213/1
A:Note: F4P12.40

Query Match 27.1%; Score 410; DB 2; Length 228;
Best Local Similarity 49.7%; Pred. No. 4.7e-23;
Matches 78; Conservative 29; Mismatches 48; Indels 2; Gaps 1;

```

```

QY 10 AGAENGSAAGGANNNGAAQQAAPAIHQDRLMPANVIRMRVLPAAHAKISDDAKETI 69
DB 2 AESQTGGGGSHESGDSQSPRLNVEQDRFLPIANVIRMRVLPAAHAKISDDAKETM 61
QY 70 QECVSEYISFTTGEANERQORRTITAEVDLWMSRLGFDYVPEPLGAYLHRYREFEG 129
DB 62 QECVSEYISFTTGEASDRKQREKRTINGDGLLWMTTLGFDYVPEPLKYLQYREMEG 121
QY 130 DARGVGLVPGAAPSRGGDHHPHSMSPAAMLSKRGVPS 166
DB 122 DTGSGG--KGESSAKRKGOPQSQVQSQVQSQVQSQGSPFS 156
RESULT 4
S22820
transcription factor NF-Y, CCAAT-binding, chain B - maize
N:Alternate names: CAAT-box DNA-binding protein
C:Species: Zea mays (maize)
C:Date: 29-Jan-1998 #sequence_revision 06-Feb-1998 #text_change 23-Jul-1999
C:Accession: S22820
R:Li, X.Y.; Mantovani, R.; van Huijsduijn, R.H.; Andre, I.; Benoist, C.; Mathis, D.
Nucleic Acids Res. 20, 1087-1091, 1992
A:Title: Evolutionary variation of the CCAAT-binding transcription factor NF-Y.
A:Reference number: S22816; MUID:92195809
A:Accession: S22820
A:Molecule type: mRNA
A:Residues: 1-179 <LIX>
A:Cross-references: EMBL:X59714
C:Superfamily: transcription factor HAP3
C:Keywords: DNA binding; transcription regulation
F:30-119/Domain: DNA binding #status predicted <DNA>

Query Match 25.8%; Score 391; DB 2; Length 179;
Best Local Similarity 44.6%; Pred. No. 8.8e-22;
Matches 82; Conservative 31; Mismatches 45; Indels 26; Gaps 5;

QY 8 PAAGA---ENGSAAGGANNNGAAQQAAPAIHQDRLMPANVIRMRVLPAAHAKISDD 64
DB 8 PGGGGGSHESGDSRGGGGG-----SVREQDRFLPIANVIRMRVLPAAHAKISDD 58
QY 65 AKETIQECVSEYISFTTGEANERQORRTITAEVDLWMSRLGFDYVPEPLGAYLHRY 124
DB 59 AKETVQECVSEYISFTTGEASDRKQREKRTINGDGLLWMTTLGFDYVPEPLKYLQY 118
QY 125 REFEGDARGVGLVPGAAPSRGGD-----HHPHSMSPAAMLSKRGVPS-CAAMLPHHHH 176
DB 119 REMEGDSK-----LTAKSSDGSIRKKDALGHVGSASSAAEGGQGGAYNQGMYNQPOYH 172
QY 177 HHDM 180
DB 173 NGDI 176

RESULT 5
A84788
Probable CCAAT-box binding transcription factor [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: A84788
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Unayam, L.; Tallon,
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487
A:Accession: A84788
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-178 <STO>
A:Cross-references: GB:AE002093; NID:g4371295; PIDN:AAD18153.1; GSPDB:GN00139
C:Genetics:
A:Gene: At2g37060

```

A;Map position: 2

[illegible]

RESULT 6
E84810
hypothetical protein At2g38880 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: E84810
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Umayam, L.; Tallon, L.;
Euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter,
Nature 402, 761-768, 1999
A:title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:reference number: A84420; MUID:20083487
A:Accession: E84810
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-138 <STO>
A:Cross-references: GB:AE002093; NID:g3928076; PIDN:AAC79602.1; GSPDB:GN00139
C:Genetics:
A:Gene: At2g38880
A:Map position: 2

Query Match	23.8%;	Score	360.5;	DB 2;	Length	138;			
Best Local Similarity	45.4%;	Pred. No.	11e-19;						
Matches	74;	Conservative	29;	Mismatches	29;	Indels	31;	Gaps	4;
QY	8	PAAGAENSGAAGANCGAAQAAPAIRQDRIMPTIANVIRMRRLPAHAKTSDDAKE	67						
Db	5	PSPAGDGGESGG-----SVREDRYLPITANISRMKKALPPNGKIGKDAKD	51						
QY	68	TIQECVSEYISFITGEANERCKQREKRTIAEDVPLWAMSLRGDFDDYVEPLGAYLHRYREF	127						
Db	52	TVQECVSEYISFITSEASDKQCKREKRTVNGDDLLWAMATLGFEDYLEPLKIYLARYR--	109						
QY	128	EGDARGVGLPGAAPSRGGDHHSPSPAAMLKSRGPVSGAAM	170						
Db	110	EGDNKGSG-----KSGD-----GSNRDAGGVSGEEM	136						

RESULT 7
G84919 probable CCAAT-box binding transcription factor [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: G84919
R.Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768. 1999

A;Title: Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*.
A;Reference number: A84420; MUID:20083487
A;Accession: G84919
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-160 <STO>
A;Cross-references: GB:AE002093; NID:g3738293; PIDN:AAC63635.1; GSPDB:GN00139
C;Genetics:
A;Gene: At2g47810
A;Map position: 2

Query Match	23.6%	Score 357.5;	DB 2;	Length 160;
Best Local Similarity	53.1%;	Pred. No. 2.e-19;		
Matches 68;	Conservative 23;	Mismatches 26;	Indels 11;	Gaps 2;
QY	28	OQHAAPATREODRLMPANTVIRIMRVLPAHAKISDDAKETMOECVSEYISFTGEANER	87	
		: :: : : : : : :		
Ddb	42	QQEESMMVKOEORLLPTIANRVGRMKNLTPANAKVSKEAKETMOECVSEFISFTGEASDK	101	
QY	88	CQOREQRKTTAEDVLWAMSRLGFGDDYVEPLGALHYRYREFEGDARGVLGPAAASRGGD	147	
		: : : : : : : :		
Ddb	102	CHKEKRKTIVNGDDICWMANLGFDYYAAQLKKYLHYRVLVEGE-----KPNNHGK	151	
QY	148	HHPHSNMP	155	
Ddb	152	GGPKS-SP	158	

RESULT 8
F84508
probable CCAAT-box binding transcription factor [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: F84508
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
Muss, D.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon,
N.; Dierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: AB84420; MUID:20083487

[illegible]

S22817

transcription factor NF-Y, CCAAT-binding, chain B - human
N:Alternate names: CAAT-box DNA-binding protein
C:Species: Homo sapiens (man)
C:Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 26-Aug-1999
C:Accession: S22817
R:Li, X.Y.; Mantovani, R.; van Huijsduijnen, R.H.; Andre, I.; Benoist, C.; Mathis, D.
Nucleic Acids Res. 20, 1087-1091, 1992

A:Title: Evolutionary variation of the CCAAT-binding transcription factor NF-Y.

A:Reference number: S22816; MUID:92195809

A:Accession: S22817

A:Molecule type: mRNA

A:Residues: 1-205 <LIA>

A:Cross-references: EMBL:X59710; NID:g35049; PIDN:CAA42230.1; PID:g35050

C:Superfamily: transcription factor HAP3

C:Keywords: DNA binding; transcription regulation

F:51-140/Domain: DNA binding #status predicted <DNA>

Query Match 21.4%; Score 324.5; DB 2; Length 205;

Best Local Similarity 56.6%; Pred. No. 7.2e-17;

Matches 64; Conservative 17; Mismatches 29; Indels 3; Gaps 1;

QY 34 AIRQDRLMPIANVIRMRVLPAAHAKISDDAKETIQECVSEYISFITGEANERCQEQR 93

: |||| : |||| ||| : || : |||| : |||| |||| |||| |||| |||| : ||

Db 49 SFREQDIYLPANVARIMKNAIPQTGKIADAKCEVCQECVSEFISFITSEASERCHQEK 108

: |||| : |||| ||| : || : |||| : |||| |||| |||| |||| |||| : ||

QY 94 KTTAEDVLWMSRLGFDYVEPLGAYLHRYREFEGDARGVGLVPGAPSRGG 146

||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||

Db 109 KTINGEDILFAMSTLGFDSYVEPLKLYLQKFRAMKGEKGIG---GAVTATDG 158

||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||

RESULT 10

F38245

transcription factor NF-Y, CCAAT-binding, chain B - mouse

C:Species: Mus musculus (house mouse)

C:Date: 28-Aug-1992 #sequence_revision 28-Aug-1992 #text_change 26-Aug-1999

C:Accession: F38245; S12045

R:Li, X.Y.; Hooft van Huijsduijnen, R.; Mantovani, R.; Benoist, C.; Mathis, D.

J. Biol. Chem. 267, 8984-8990, 1992

A:Title: Intron-exon organization of the NF-Y genes. Tissue-specific splicing modifies a

A:Reference number: A38245; MUID:92250488

A:Accession: F38245

A:Status: preliminary; not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 1-207 <LIA>

A:Cross-references: GB:M86215

R:Hooft van Huijsduijnen, R.; Li, X.Y.; Black, D.; Matthes, H.; Benoist, C.; Mathis, D.

EMBO J. 9, 3119-3127, 1990

A:Title: Co-evolution from yeast to mouse: cDNA cloning of the two NF-Y (CP-1/CBF) subun

A:Reference number: S12044; MUID:91006004

A:Accession: S12045

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-207 <HOO>

A:Cross-references: GB:X5316; NID:g53362; PIDN:CAA39024.1; PID:g53363

C:Superfamily: transcription factor HAP3

C:Keywords: alternative splicing; DNA binding; transcription regulation

Query Match

Best Local Similarity 56.6%; Pred. No. 7.3e-17;

Matches 64; Conservative 17; Mismatches 29; Indels 3; Gaps 1;

QY 34 AIRQDRLMPIANVIRMRVLPAAHAKISDDAKETIQECVSEYISFITGEANERCQEQR 93

: |||| : |||| ||| : || : |||| : |||| |||| |||| |||| |||| : ||

Db 51 SFREQDIYLPANVARIMKNAIPQTGKIADAKCEVCQECVSEFISFITSEASERCHQEK 110

: |||| : |||| ||| : || : |||| : |||| |||| |||| |||| |||| : ||

QY 94 KTTAEDVLWMSRLGFDYVEPLGAYLHRYREFEGDARGVGLVPGAPSRGG 146

||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||

Db 111 KTINGEDILFAMSTLGFDSYVEPLKLYLQKFRAMKGEKGIG---GAVTATDG 160

||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||

RESULT 11

A23692

transcription factor, CCAAT-binding, chain A1 - rat

C:Species: Rattus norvegicus (Norway rat)

C:Date: 04-Oct-1991 #sequence_revision 04-Oct-1991 #text_change 26-Aug-1999

C:Accession: A23692

R:Vuorio, T.; Maity, S.N.; de Crombrughe, B.

J. Biol. Chem. 265, 22480-22486, 1990

A:Title: Purification and molecular cloning of the "A" chain of a rat heteromeric CCA

A:Reference number: A23692; MUID:91093096

A:Accession: A23692

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-207 <VUO>

A:Cross-references: GB:M5045; GB:J05701; NID:g203352; PIDN:AAA40887.1; PID:g203353

C:Superfamily: transcription factor HAP3

C:Keywords: alternative splicing; DNA binding; transcription regulation

F:53-142/Domain: DNA binding #status predicted <DNA>

Query Match 21.4%; Score 324.5; DB 2; Length 207;

Best Local Similarity 56.6%; Pred. No. 7.3e-17;

Matches 64; Conservative 17; Mismatches 29; Indels 3; Gaps 1;

QY 34 AIRQDRLMPIANVIRMRVLPAAHAKISDDAKETIQECVSEYISFITGEANERCQEQR 93

: |||| : |||| ||| : || : |||| : |||| |||| |||| |||| |||| : ||

Db 51 SFREQDIYLPANVARIMKNAIPQTGKIADAKCEVCQECVSEFISFITSEASERCHQEK 110

: |||| : |||| ||| : || : |||| : |||| |||| |||| |||| |||| : ||

QY 94 KTTAEDVLWMSRLGFDYVEPLGAYLHRYREFEGDARGVGLVPGAPSRGG 146

||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||

Db 111 KTINGEDILFAMSTLGFDSYVEPLKLYLQKFRAMKGEKGIG---GAVSATDG 160

||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||

RESULT 12

S24469

transcription factor NF-Y, CCAAT-binding, chain B - chicken

C:Species: Gallus gallus (chicken)

C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 26-Aug-1999

C:Accession: S24469

R:Benoist, C.

submitted to the EMBL Data Library, January 1992

A:Reference number: S24469

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-180 <BEN>

A:Cross-references: EMBL:X59713; NID:g63690; PIDN:CAA42233.1; PID:g63691

C:Superfamily: transcription factor HAP3

C:Keywords: DNA binding; transcription regulation

F:51-140/Domain: DNA binding #status predicted <DNA>

Query Match

Best Local Similarity 59.8%; Pred. No. 9.5e-17;

Matches 61; Conservative 16; Mismatches 25; Indels 0; Gaps 0;

QY 34 AIRQDRLMPIANVIRMRVLPAAHAKISDDAKETIQECVSEYISFITGEANERCQEQR 93

: |||| : |||| ||| : || : |||| : |||| |||| |||| |||| |||| : ||

Db 49 SFREQDIYLPANVARIMKNAIPQTGKIADAKCEVCQECVSEFISFITSEASERCHQEK 108

: |||| : |||| ||| : || : |||| : |||| |||| |||| |||| |||| : ||

QY 94 KTTAEDVLWMSRLGFDYVEPLGAYLHRYREFEGDARGVG 135

||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||

Db 109 KTINGEDILFAMSTLGFDSYVEPLKLYLQKFRAMKGEKGIG 150

||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||

RESULT 13

JC6080

transcription factor HAP3 - Emericella nidulans

C:Species: Emericella nidulans, Aspergillus nidulans

C:Date: 21-Jan-1997 #sequence_revision 21-Jan-1997 #text_change 26-Aug-1999

C:Accession: JC6080

R:Papagiannopoulos, P.; Andrianopoulos, A.; Sharp, J.A.; Davis, M.A.; Hynes, M.J.

Mol. Gen. Genet. 251, 412-421, 1996

A:Title: The hapC gene of *Aspergillus nidulans* is involved in the expression of CCAAT-CD
A:Reference number: JC6080; MUID:96285564

A:Accession: JC6080

A:Molecule type: mRNA

A:Residues: 1-186 <PAP>

A:Cross-references: GB:U35341; NID:g1017715; PIDN:AAC49411.1; PID:g1017716

C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001

C:Accession: C86222

R:Theologos, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon

Chen, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, T.H.; Dewar,

ausen, N.F.; Hughes, B.; Huizar, L.

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzia

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant *Arabidopsis*.

A:Reference number: A86141; MUID:21016719

A:Accession: C86222

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-139 <STO>

A:Cross-references: GB:AE005172; NID:g1922961; PIDN:AAE70405.1; GSPDB:GN00141

C:Genetics:

A:Map position: 1

Query Match 21.1%; Score 320; DB 2; Length 186;

Best Local Similarity 53.6%; Pred. No. 1.4e-16;

Matches 67; Conservative 25; Mismatches 29; Indels 4; Gaps 3;

QY 35 IREODRLMPIANVIRIMRRVLPAAKISDDAKETIOECVSEYISFITGEANERCQREQRK 94

DB 41 VKEDRWLPANVARIMKALPENAKIAKECMQECVSEFISFITSEASEKCOQEKRR 100

QY 95 TITAEVDLWMSRLGFDYVEPLGAYLHRYREPEGDARGVGL-VPGAAPSRGGDHPHSM 153

DB 101 TVNGEDILFAMTSLGFENYAEALKIYLSKYRETQ-SARGEHQNRPPSSGYAGGD--PSAE 157

QY 154 SPRAAM 158

DB 158 LPAAL 162

RESULT 14

S22818

transcription factor NF-Y, CCAAT-binding, chain B - sea lamprey

N:Alternate names: CCAAT-box DNA-binding protein

C:Species: *Petromyzon marinus* (sea lamprey)

C:Date: 29-Jan-1998 #sequence_revision 06-Feb-1998 #text_change 26-Aug-1999

C:Accession: S22818; S78116

R:Li, X.Y.; Mantovan, R.; van Huijsduijn, R.H.; Andre, I.; Benoist, C.; Mathis, D.

Nucleic Acids Res. 20, 1087-1091, 1992

A:Title: Evolutionary variation of the CCAAT-binding transcription factor NF-Y.

A:Reference number: S22816; MUID:92195809

A:Accession: S22818

A:Molecule type: mRNA

A:Residues: 1-209 <LIX>

A:Cross-references: EMBL:X59712

R:Benoist, C.

submitted to the EMBL Data Library, January 1992

A:Reference number: S78116

A:Accession: S78116

A:Molecule type: mRNA

A:Residues: 1-110, R', 112-209 <BEN>

A:Cross-references: EMBL:X59712; NID:g64217; PIDN:CAA42232.1; PID:g64218

C:Superfamily: transcription factor HAP3

C:Keywords: DNA binding; transcription regulation

F:54-143/Domain: DNA binding #status predicted <DNA>

Query Match 21.1%; Score 319.5; DB 2; Length 209;

Best Local Similarity 52.4%; Pred. No. 1.7e-16;

Matches 65; Conservative 18; Mismatches 38; Indels 3; Gaps 1;

QY 11 GAEGSLASGDHDECGSK---DPREQDIYLPANVARIMKTSIPSSGKIADAKECVQ 70

DB 32 GAEGSLASGDHDECGSK---DPREQDIYLPANVARIMKTSIPSSGKIADAKECVQ 88

QY 71 ECVSEYISFITGEANERCQREQRKTITAEVDLWMSRLGFDYVEPLGAYLHRYREPEGD 130

DB 89 ECVSEYISFITGEANERCQREQRKTITAEVDLWMSRLGFDYVEPLGAYLHRYREPEGD 130

QY 131 ARGV 134

DB 149 ESKI 152

Result No.	Query No.	Score	Query		Length	DB	ID	Description
			Match	%				
1	1	499.5	33.0	205	10	Q9FGJ0	Q9fgj0 arabidopsis	
2	2	494	32.6	208	10	Q8L130	Q8ll30 arabidopsis	
3	3	494	32.6	208	10	Q9SFD8	Q9sfd8 arabidopsis	
4	4	443.5	29.3	187	10	Q23634	Q23634 arabidopsis	
5	5	443.5	29.3	190	10	Q9FGJ3	Q9fgj3 arabidopsis	
6	6	410.5	27.1	161	10	Q23310	Q23310 arabidopsis	
7	7	410	27.1	228	10	Q9LFR3	Q9lfr3 arabidopsis	
8	8	376.5	24.9	141	10	Q23633	Q23633 arabidopsis	
9	9	372	24.6	178	10	Q9ZQC3	Q9zqc3 arabidopsis	
10	10	360.5	23.8	138	10	Q9SLG0	Q9slg0 arabidopsis	
11	11	357.5	23.6	160	10	Q82248	Q82248 arabidopsis	
12	12	354.5	23.4	215	10	Q9SIT9	Q9sit9 arabidopsis	
13	13	344.5	22.8	215	3	Q59848	Q59848 aspergillus	
14	14	331.5	21.9	206	13	Q73744	Q73744 xenopus lae	
15	15	324.5	21.4	148	11	Q63091	Q63091 rattus norv	
16	16	320	21.1	186	3	Q60735	Q60735 emericea	
17	17	313.5	20.7	242	5	Q76236	Q76236 schistosoma	
18	18	307	20.3	139	10	Q04027	Q04027 arabidopsis	
19	19	289.5	19.1	183	5	Q9V156	Q9v156 drosophila	

```

081130
ID O81130 PRELIMINARY; PRT; 208 AA.
AC O81130;
DT 01-NOV-1998 (TReMBLrel. 08, Created)
DT 01-NOV-1998 (TReMBLrel. 08, Last sequence update)
DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)
DE CCAAT-BOX BINDING FACTOR HAP3 HOMOLOG.
OS Arabidopsis thaliana (Mouse-ear cross).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WS-0;
RX MEDLINE=98319234; PubMed=9657152;
RA Lotan T., Ohto T., Yee K., West M.A.L., Lo R., Kwong R.M.,
RA Yamagishi K., Fischer R.L., Goldberg R.B., Hatada J.J.;
RT "Arabidopsis LEAFY COTYLEDON1 is sufficient to induce embryo
RT development in vegetative cells.";
RL Cell 93:1195-1205(1998).
DR EMBL; AF036684; AAC39488.1; -.
DR HSP; P19267; IB6W.
DR InterPro; IPR00166; -.
DR InterPro; IPR000947; -.
DR Pfam; PF00808; CBFD_NFYB_HMF; 1.
DR PRINTS; PR00615; CCAATSUBUNTA.
DR SEQUENCE 208 AA; 22693 MW; 4E2D249AE2525DDB CRC64;

Query Match 32.6%; Score 494; DB 10; Length 208;
Best Local Similarity 47.0%; Pred. No. 1.5e-32;
Matches 117; Conservative 19; Mismatches 47; Indels 66; Gaps 9;

QY 16 SAAGGANNGAAQQAAPAIREQDRLMPIANVIRIMRRVLPAAHAKISDDAKETIQCVSE 75
: || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 8 AGAGDKNNGIVVQQPPCVAREQDQYMPIANVIRIMRKTLPASHAKISDDAKETIQCVSE 67

QY 76 YISFTIGEANERCQQRKTITAEIDLWMSRLGFDYVEPLGAYLHRYREFEGDARGV 135
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 68 YISFTIGEANERCQQRKTITAEIDLWMSRLGFDYVEPLGAYLHRYREFETD-RGSA 126

QY 136 LVPGAAPSRGGDHHPHSMSPAAMLKSRGPVSGAAMLPHHHHHHMQMHAAMYGCTAVPPP 195
|| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 127 L-----RG-----EPPSLRQT-----YGGNGI--- 143

QY 196 AGPPHHGGFLMPHPQGS SHYL----PYAYEPTYGGEHAMAAYYGAAYAPNGSGSGDG 251
|| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 144 -----GF-----HGPSHGLPPPGPGY-----GMLDQSMVMGGRYYQ-NGSSGQDES 185

QY 252 SGGGGGSAS 260
|| ||| |||
Db 186 SVGGSSSS 194

RESULT 3
Q9SFD8 PRELIMINARY; PRT; 208 AA.
ID Q9SFD8;
AC Q9SFD8;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TReMBLrel. 13, Last annotation update)
DE T26F17.20.
OS Arabidopsis thaliana (Mouse-ear cross).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Shinn P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C., Khan S.,
RA Kim C., Altafi H., Bei Q., Chin C., Chioi J., Choi E., Conn L.,
RA Conway A., Gonzales A., Hansen N., Howing B., Koo T., Lam B., Lee J.,

```

```

RA Lenz C., Li J., Liu A., Liu K., Liu S., Mukharsky N., Nguyen M.,
RA Palm C., Pham P., Sakano H., Schwartz J., Southwick A., Thaveri A.,
RA Toriumi M., Vaysberg M., Yu G., Federspiel N.A., Theologis A.,
RA Ecker J.R.;
RT "Genomic sequence for Arabidopsis thaliana BAC T36F17 from chromosome
RT 1.";
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC013482; AAF16537.1; -.
DR HSP; P19267; IB6W.
DR InterPro; IPR00166; -.
DR InterPro; IPR000947; -.
DR Pfam; PF00808; CBFD_NFYB_HMF; 1.
DR PRINTS; PR00615; CCAATSUBUNTA.
DR SEQUENCE 208 AA; 22679 MW; 0DEB469D9F8BEFDC CRC64;

Query Match 32.6%; Score 494; DB 10; Length 208;
Best Local Similarity 47.0%; Pred. No. 1.5e-32;
Matches 117; Conservative 19; Mismatches 47; Indels 66; Gaps 9;

QY 16 SAAGGANNGAAQQAAPAIREQDRLMPIANVIRIMRRVLPAAHAKISDDAKETIQCVSE 75
: || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 8 AGAGDKNNGIVVQQPPCVAREQDQYMPIANVIRIMRKTLPASHAKISDDAKETIQCVSE 67

QY 76 YISFTIGEANERCQQRKTITAEIDLWMSRLGFDYVEPLGAYLHRYREFEGDARGV 135
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 68 YISFTIGEANERCQQRKTITAEIDLWMSRLGFDYVEPLGAYLHRYREFETD-RGSA 126

QY 136 LVPGAAPSRGGDHHPHSMSPAAMLKSRGPVSGAAMLPHHHHHHMQMHAAMYGCTAVPPP 195
|| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 127 L-----RG-----EPPSLRQT-----YGGNGI--- 143

QY 196 AGPPHHGGFLMPHPQGS SHYL----PYAYEPTYGGEHAMAAYYGAAYAPNGSGSGDG 251
|| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 144 -----GF-----HGPSHGLPPPGPGY-----GMLDQSMVMGGRYYQ-NGSSGQDES 185

QY 252 SGGGGGSAS 260
|| ||| |||
Db 186 SVGGSSSS 194

RESULT 4
O23634 PRELIMINARY; PRT; 187 AA.
ID O23634;
AC O23634;
DT 01-JAN-1998 (TReMBLrel. 05, Created)
DT 01-JAN-1998 (TReMBLrel. 05, Last sequence update)
DT 01-OCT-2000 (TReMBLrel. 15, Last annotation update)
DE TRANSCRIPTION FACTOR (FRAGMENT).
GN HAP3B.
OS Arabidopsis thaliana (Mouse-ear cross).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Edwards D., Smith A.G., Murray J.A.;
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; Y13724; CAAT4052.1; -.
DR Mendel; 26811; Arath; 2691; 26811.
DR InterPro; IPR000166; -.
DR InterPro; IPR000947; -.
DR Pfam; PF00808; CBFD_NFYB_HMF; 1.
DR PRINTS; PR00615; CCAATSUBUNTA.
DR PROSITE; PS00685; CBFA_NFYB; 1.
FT NON_TER 1
SQ SEQUENCE 187 AA; 20295 MW; E4C7ABC89F37F40E CRC64;

Query Match 29.3%; Score 443.5; DB 10; Length 187;
Best Local Similarity 38.7%; Pred. No. 1.6e-28;
Matches 94; Conservative 29; Mismatches 55; Indels 65; Gaps 3;

```

IC	023310	PRELIMINARY;	PRT;	161 AA.
AD	023310;			
DT	01-JAN-1998	(TREMBLrel. 05, Created)		
DT	01-JAN-1998	(TREMBLrel. 05, Last sequence update)		
DT	01-JAN-2001	(TREMBLrel. 16, Last annotation update)		
DT	01-WAR-2001	(TREMBLrel. 16, Last annotation update)		
DT	CCAT-BINDING	TRANSCRIPTION FACTOR SUBUNIT A.		
DN	AT4G14540.			
GN	Arabidopsis thaliana (Mouse-ear cress).			
OC	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;			
OC	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;			
OC	Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;			
OC	Brassicales; Brassicaceae; Arabidopsis.			
ON	NCBI_TaxID=3702;			
OX	[1]			
RP	SEQUENCE FROM N.A.			
RA	Bevan M., Stickema W., Murphy G., Wambutt R., Pohl T., Terryn N.,			
RA	Kreis M., Kavanagh T., Entlan K.D., Rieger M., James R.,			
RA	Puidgomech P., Hatzopoulos P., Obermaier B., Duesterhoft A.,			
RA	Jones J., Palme K., Ansorge W., Pilsen M., Bancroft I., Mewes H.W.			
RA	Schuelker C., Chaiwatiz N.,			
RL	Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.			

```

RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: Z97336; CAB10233.1; -.
DR EMBL: AL161539; CAB78496.1; -.
DR HSSP: PI9267; 1B6W.
DR Mendel: 26681; Arath:2691;26681.
DR InterPro: IPR000166; -.
DR InterPro: IPR000947; -.
DR Pfam: PF0808; CBFD_NFYB_HMF; 1.
DR PRINTS: PR00615; CCAATSUBUNTA.
DR PROSITE: PS00685; CBFA_NFYB; 1.
DR SEQUENCE 161 AA; 17186 MW; 5C452E20D98A7AF4 CRC64;

Query Match 27.1%; Score 410.5; DB 10; Length 161;
Best Local Similarity 48.3%; Pred. No. 6.2e-26;
Matches 87; Conservative 23; Mismatches 35; Gaps 3;

QY 12 AENGSAAGGANNCGAAQQAAPAREODRLMPTIANVIRIMRVLPAAKISDADKETIOE 7
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 2 ADSNDSSGGHGKGGNA-----STREQDRFLPIANVSRIMKKALPANAKISKADKETVOE 5

```

Db	56	CVSEISFITGEASDKCQREKRKTINGDDLWANTTLGFEDYVEPLKVYLQKYREVEGEK
Qv	129	-----GDARGVGLVPGAAPSRGGDHHHPHSPSPAAMLKSRGPVSGAAM---LPHHHHHH

Db	116	TTTTAGROGDKEGGGGGAGSGG-----APMYGGGVTTMGHQFSHH	
RESULT	7		
Q9LFI3			
ID	Q9LFI3	PRELIMINARY;	PRT; 228 AA.
AC	Q9LFI3;		
DT	01-OCT-2000	(T-EMBLrel. 15, Created)	
DT	01-OCT-2000	(T-EMBLrel. 15, Last sequence update)	
DT	01-MAR-2001	(T-EMBLrel. 16, Last annotation update)	
DE	DE TRANSCRIPTION FACTOR NF-Y, CCAAT-BINDING-LIKE PROTEIN.		
GN	F4P12.40.		
OS	Arabidopsis thaliana (Mouse-ear cress);		
OC	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta		
OC	Magnoliophyta; eudicotyledons; core eudicots; rosidae; eurosids II		
OC	Brassicales; Brassicaceae; Arabidopsi.		
OX	NCBI_TaxID=3702;		
RN	[1]		
RP	SEQUENCE FROM N.A.		

Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.

```

Db      112 EGNKGGSG-----KSGD-----GSMRAGGVGSGEEM 138
          :|||::| | |   : : : ||||| |
RESULT    9
Q9ZQC3 ID Q9ZQC3 PRELIMINARY; PRT; 178 AA.
AC Q9ZQC3;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DDT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DE 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DEFUTATIVE CCAAT-BINDING TRANSCRIPTION FACTOR.
GN T2N18.18.
OS Arabidopsis thaliana (Mouse-ear cross).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsids.
ON NCBI_TaxID=3702;
OX [1]
PC SEQUENCE FROM N.A.
RP STRAIN=Cv. COLUMBIA;
RA Lin X., Kaul S., Shea T.P., Fujii C.Y., Shen M., VanAken S.E.,
RA Barstead M.E., Mason T.M., Bowman C.L., Ronning C.M., Benito M.,
RA Carrera A.J., Creasy T.H., Buell C.R., Town C.D., Niezman W.C.,
RA Fraser C.M., Venter J.C.;
RA "Arabidopsis thaliana chromosome II BAC T2N18 genomic sequence.";
RL Submitted (MAR-1999) to the EMBL/GenBank/DDBJ databases.
RD EMBL; AC006260; AAC18153.1; -
DR Mendel; 39697; Arath; 2691; 39697.
DR InterPro; IPR000166; -.
DR InterPro; IPR000947; -.
DR Pfam; PF00808; CBFD_NFYB_HMF; 1.
DR PRINTS; PR00615; CCAATSUBUNTA.
DR PROSITE; PS00685; CBFA_NFYB; 1.
SQ SEQUENCE 178 AA; 19523 MW; B3CF497383EC86C4 CRC64;

Query Match 24.6%; Score 372; DB 10; Length 178;
Best Local Similarity 43.6%; Pred. No. 9e-23;
Matches 79; Conservative 27; Mismatches 49; Indels 26; Gaps

QY 20 GANNCGAAQQHAAPAIREDQLMPIANVIRMRVLPAHAKTSDDAKETIQCVSEYISF 79
       |:| | | | | :|:|||||:|:| | | | | | | | | | | | | | | | | |
Db 13 GSHEGGGDSPRSUHVREQRDFLPANISRIMKRGLPANGKIADKAKEIVQEVSFI 72
       |TGAEANCRCQRKTITAEEDVLWMSRLGFDDYVEPLGAYLHYREFEGDARGVLVP 139
       73 VTSEASDKCRKERTKGNDLLWANATIGFDYEHPVKVLMRY--EGDTKG----- 124
Db 73 VTSSEASDKCRKERTKGNDLLWANATIGFDYEHPVKVLMRY--EGDTKG----- 124

QY 140 AAPSRGGDHPHSWSPAAMLSRGPSVGSAAMLPHHHHHMDMHMAAMYGTAVPPAGPP 199
       ::||| | | | | | | | | | | | | | | | | | | | | | | | |
Db 125 --SARKGD-----PNNAKKDGSSQG-----QFSOLAHQGPYGNSQVTFLFS 166
       | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 200 H 200
Db 167 H 167

RESULT 10
Q9SLGO ID Q9SLGO PRELIMINARY; PRT; 138 AA.
AC Q9SLGO;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DDT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DE 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DEFUTATIVE CCAAT-BINDING TRANSCRIPTION FACTOR SUBUNIT.
GN AT2G3880.
OS Arabidopsis thaliana (Mouse-ear cross).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsids.
ON NCBI_TaxID=3702;
OX [1]

```

```
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RX MEDLINE=20083487; PubMed=10617197;
RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,
RA Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,
RA Cronin L.A., Shen M., VanAken S.E., Umayam L., Tallon L.J., Gill J.E.,
RA Adams M.D., Carrera A.J., Cressy T.H., Goodman H.M., Somerville C.R.,
RA Copenhaver G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,
RA Salzberg S.L., Fraser C.M., Venter J.C.;
RT "Sequence and analysis of chromosome II of Arabidopsis thaliana.";
RL Nature 402:761-768(1999).
DR EMBL: AC005770; AAC79602.1; -.
DR HSSP: P19267; 1B6W.
DR InterPro: IPR000166; -.
DR InterPro: IPR000947; -.
DR Pfam: PF00808; CBFD_NFYB_HMF; 1.
DR PRINTS: PR00615; CCAATSUBUNTA.
DR PROSITE: PS00685; CBFA_NFYB; 1.
SQ SEQUENCE 138 AA; 14764 MW; 14CE11FC21D5768A CRC64;

Query Match 23.8%; Score 360.5; DB 10; Length 138;
Best Local Similarity 45.8%; Pred. No. 5.7e-22;
Matches 74; Conservative 29; Mismatches 29; Indels 31; Gaps 4;

QY 8 PAAGAENGSAAGNAGAAQAQAAPAIREDQRLMPIANVIRIMRRLPAHAHAKISDDAKE 67
DB 5 PSSPAGDGGESG-----SVREQDRLYLANISIRIMKALPPNGKIGKDKAD 51

QY 68 TIOECVSEYISFTIGANERQORQRTITAEDVLWMSRLGFDVVEPLGAYLHRYREF 127
DB 52 TVQECVSEYISFTSEASDKQCKRKTNGDGLLWAMATLGFEDYLEPLKIYLYR-- 109

QY 128 EGDARGVGLVGAAPRGDGHHPHSPAMLSKRGCPVSSAAM 170
DB 110 EGNKGG-----KSGD-----GSDRAGGGVSGEEM 136

RESULT 11
ID 082248 PRELIMINARY; PRT; 160 AA.
AC 082248;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE PUTATIVE CCAAT-BINDING TRANSCRIPTION FACTOR SUBUNIT A (CBF-A).
GN F17A22.20.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RX MEDLINE=20083487; PubMed=10617197;
RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,
RA Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,
RA Cronin L.A., Shen M., VanAken S.E., Umayam L., Tallon L.J., Gill J.E.,
RA Adams M.D., Carrera A.J., Cressy T.H., Goodman H.M., Somerville C.R.,
RA Copenhaver G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,
RA Salzberg S.L., Fraser C.M., Venter J.C.;
RT "Sequence and analysis of chromosome II of Arabidopsis thaliana.";
RL Nature 402:761-768(1999).
DR EMBL: AC007063; AAD22680.1; -.
DR HSSP: P48781; 1B67.
DR InterPro: IPR000166; -.
DR InterPro: IPR000947; -.
DR Pfam: PF00808; CBFD_NFYB_HMF; 1.
DR PRINTS: PR00615; CCAATSUBUNTA.
DR PROSITE: PS00685; CBFA_NFYB; 1.
SQ SEQUENCE 215 AA; 24619 MW; 0615B2B7A097DD25 CRC64;

Query Match 23.4%; Score 354.5; DB 10; Length 215;
Best Local Similarity 37.8%; Pred. No. 2.9e-21;
Matches 79; Conservative 31; Mismatches 54; Indels 45; Gaps 6;

QY 8 PAAGAEN-GSAAGNAGAAQAQAAPAIREDQRLMPIANVIRIMRRLPAHAHAKISDDAK 66
DB 13 PGVAETNPGSPSSKNTNNNNN-----KEQDRELPANVGRIMKKVLPNGKISKDAK 65

QY 67 ETIOECVSEYISFTIGANERQORQRTITAEDVLWMSRLGFDVVEPLGAYLHRYRE 126
DB 66 ETVOECVSEYISFTIGANERQORQRTITAEDVLWMSRLGFDVVEPLGAYLHRYRE 125

QY 127 FEGDARGVGLVGAAPS-----RGDHPHSPAMLSKRGCPVSSAAMLPHHH 175
DB 126 TEGEKVN-----SPKQQQQQQQQQQQQQHHYQEQED--QNNNNMSCTSISSHH 176

QY 176 HHMDMOMHAAMYGGTAVPPPPAGPHHGGF 204
DB 177 -----PSPFLPDHQP 188

RESULT 13
```

```
QY 28 QOHAAPAIREDQRLMPIANVIRIMRRLPAHAHAKISDDAKETIOECVSEYISFTIGANER 87
DB 42 QOESMMVKEQDRLMPIANVGRIMKNILPANAKYSKEAKETMQECVSEYISFTVGEASDK 101

QY 88 CQORQRTITAEDVLWMSRLGFDVVEPLGAYLHRYREFEGDARGVGLVGAAPSRRGD 147
DB 102 CHKEKRTVNGDDICWAMANLGFDDYAAQLKKYLHRYRLEGE-----KPNHGGK 151

QY 148 HPHMSMP 155
DB 152 GPKS-SP 158

RESULT 12
ID 09SIT9 PRELIMINARY; PRT; 215 AA.
AC 09SIT9;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE PUTATIVE CCAAT-BOX BINDING TRANSCRIPTION FACTOR.
GN AT2G13570.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RX MEDLINE=20083487; PubMed=10617197;
RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,
RA Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,
RA Cronin L.A., Shen M., VanAken S.E., Umayam L., Tallon L.J., Gill J.E.,
RA Adams M.D., Carrera A.J., Cressy T.H., Goodman H.M., Somerville C.R.,
RA Copenhaver G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,
RA Salzberg S.L., Fraser C.M., Venter J.C.;
RT "Sequence and analysis of chromosome II of Arabidopsis thaliana.";
RL Nature 402:761-768(1999).
DR EMBL: AC007063; AAD22680.1; -.
DR HSSP: P48781; 1B67.
DR InterPro: IPR000166; -.
DR InterPro: IPR000947; -.
DR Pfam: PF00808; CBFD_NFYB_HMF; 1.
DR PRINTS: PR00615; CCAATSUBUNTA.
DR PROSITE: PS00685; CBFA_NFYB; 1.
SQ SEQUENCE 215 AA; 24619 MW; 0615B2B7A097DD25 CRC64;

Query Match 23.4%; Score 354.5; DB 10; Length 215;
Best Local Similarity 37.8%; Pred. No. 2.9e-21;
Matches 79; Conservative 31; Mismatches 54; Indels 45; Gaps 6;

QY 8 PAAGAEN-GSAAGNAGAAQAQAAPAIREDQRLMPIANVIRIMRRLPAHAHAKISDDAK 66
DB 13 PGVAETNPGSPSSKNTNNNNN-----KEQDRELPANVGRIMKKVLPNGKISKDAK 65

QY 67 ETIOECVSEYISFTIGANERQORQRTITAEDVLWMSRLGFDVVEPLGAYLHRYRE 126
DB 66 ETVOECVSEYISFTIGANERQORQRTITAEDVLWMSRLGFDVVEPLGAYLHRYRE 125

QY 127 FEGDARGVGLVGAAPS-----RGDHPHSPAMLSKRGCPVSSAAMLPHHH 175
DB 126 TEGEKVN-----SPKQQQQQQQQQQQQQHHYQEQED--QNNNNMSCTSISSHH 176

QY 176 HHMDMOMHAAMYGGTAVPPPPAGPHHGGF 204
DB 177 -----PSPFLPDHQP 188

RESULT 13
```

```
O59848
ID O59848 PRELIMINARY; PRT; 215 AA.
AC O59848;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE HAPC.
GN HAPC.
OS Aspergillus oryzae.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eukariotales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX NCBI_TaxID=5062;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JCM02238;
RA Kato M., Kobayashi T., Tsukagoshi N., Tanaka A., Hashimoto H.;
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
DE EMBL; AB010431; BAA28356.1;
DR HSSP; P48781; 1B67.
DR InterPro; IPR000166;
DR InterPro; IPR000947;
DR Pfam; PF00808; CBFD_NFYB_HMF; 1.
DR PRINTS; PR00615; CCAATSUBUNTA.
DR PROSITE; PS00685; CBFA_NFYB; 1.
SQ SEQUENCE 215 AA; 23348 MW; 7A1A74E8F2E05C06 CRC64;

Query Match 22.8%; Score 344.5; DB 3; Length 215;
Best Local Similarity 38.5%; Pred. No. 1.9e-20;
Matches 77; Conservative 28; Mismatches 48; Indels 47; Gaps 4;

QY 27 AQHAAPATREQRLPIANVIRIMRRVLPAAHAKISDDAKETIQECVSEYISFITGEANE 86
DB 33 AQOGEFEVKEQRLPIANVIRIMRRVLPAAHAKISDDAKETIQECVSEYISFITGEANE 92
QY 87 RCQREKRTTADVLWMSRLGDDYVEPLGAYLHRYREFE----- 128
DB 93 KCOEKRKTNGEDILFAMTSLGFGFYAALYLSKYRETOSARGERQNRRTSSGYASG 152
QY 129 GDARGVGLVPGAAPSRGG-----DHHPHSMSPAAMLKSRGPVSGAALPHHHHHHMQM 182
DB 153 GPVGGVSSAPGGRPATAGFPDAADNTNSIMPDLPTQDP----- 194
QY 183 HAAMYGTVAPPPAGPPHUG 202
DB 195 --SAYG---YPPWVGQPHNG 209

RESULT 14
ID O73744 PRELIMINARY; PRT; 206 AA.
AC O73744;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE NUCLEAR Y/CCAAT-BOX BINDING FACTOR B SUBUNIT NF-YB.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RA Herlihy M., Wolfe A.P.;
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
DE EMBL; AF041204; AAC82336.1;
DR HSSP; P48781; 1B67.
DR InterPro; IPR000166;
DR InterPro; IPR000947;
DR Pfam; PF00808; CBFD_NFYB_HMF; 1.
DR PRINTS; PR00615; CCAATSUBUNTA.
DR PROSITE; PS00685; CBFA_NFYB; 1.
SQ SEQUENCE 206 AA; 22582 MW; 359333B20A6D43607 CRC64;
```

```
Query Match 21.9%; Score 331.5; DB 13; Length 206;
Best Local Similarity 51.9%; Pred. No. 2e-19;
Matches 69; Conservative 19; Mismatches 42; Indels 3; Gaps 2;

QY 34 AIREQRLMPIANVIRIMRRVLPAAHAKISDDAKETIQECVSEYISFITGEANERCQREQ 93
DB 50 SFREQDIYLPANVARIMKNVAPQTKIAKDAKEQCVSEFISFITSEASERCHQEK 109
QY 94 KTIATADVWMSRLGDDYVEPLGAYLHRYREFEGDARGVGLVPGAAPSRGG--HHPH 151
DB 110 KTINGEDILFAMSLGDDYVEPLKLYLQKPREAMKGEKIGGVTTTGDGLGEDLAEPF 169
QY 152 -SMSPAAMLKSRG 163
DB 170 TSQIPAGLITTDG 182
```

```
RESULT 15
ID O63091 PRELIMINARY; PRT; 148 AA.
AC O63091;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE CCAAT BINDING TRANSCRIPTION FACTOR-B SUBUNIT.
GN CBF-A11.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=91093096; PubMed=2266139;
RT Vuorio T., Maiti S.N., de Crombrughe B.;
RT "Purification and molecular cloning of the 'A' chain of a rat
RT heteromeric CCAAT-binding protein. Sequence identity with the yeast
RT HAP3 transcription factor."
RL J. Biol. Chem. 265:22480-22486(1990).
DR EMBL; M60617; AAA40888.1;
DR HSSP; P48781; 1B67.
DR InterPro; IPR000166;
DR InterPro; IPR000947;
DR Pfam; PF00808; CBFD_NFYB_HMF; 1.
DR PRINTS; PR00615; CCAATSUBUNTA.
DR PROSITE; PS00685; CBFA_NFYB; 1.
SQ SEQUENCE 148 AA; 16302 MW; 6855E5C857115AD0 CRC64;
```

```
Query Match 21.4%; Score 324.5; DB 11; Length 148;
Best Local Similarity 56.6%; Pred. No. 5e-19;
Matches 64; Conservative 17; Mismatches 29; Indels 3; Gaps 1;

QY 34 AIREQRLMPIANVIRIMRRVLPAAHAKISDDAKETIQECVSEYISFITGEANERCQREQ 93
DB 13 SFREQDIYLPANVARIMKNVAPQTKIAKDAKEQCVSEFISFITSEASERCHQEK 72
QY 94 KTIATADVWMSRLGDDYVEPLGAYLHRYREFEGDARGVGLVPGAAPSRGG 146
DB 73 KTINGEDILFAMSLGDDYVEPLKLYLQKPREAMKGEKIG--GAVSATDG 122
```

Search completed: October 25, 2001, 09:24:34
Job time: 189 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 25, 2001, 10:40:20 ; Search time 994.71 Seconds
(without alignments)
11147.166 Million cell updates/sec

Title: US-09-435-054-1
Perfect score: 1173
Sequence: 1 ccacgcgtccgccaccacac.....tcaaaaaaaaaaaaaaa 1173

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues
Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST: *
1: gb_est1.*
2: gb_est2.*
3: gb_est3.*
4: gb_est4.*
5: gb_est5.*
6: gb_est6.*
7: gb_est7.*
8: gb_est8.*
9: gb_est9.*
10: gb_est10.*
11: gb_est11.*
12: gb_est12.*
13: gb_est13.*
14: gb_est14.*
15: gb_est15.*
16: gb_est16.*
17: gb_est17.*
18: gb_est18.*
19: gb_est19.*
20: gb_est20.*
21: gb_est21.*
22: gb_est22.*
23: gb_est23.*
24: gb_est24.*
25: gb_est25.*
26: gb_est26.*
27: gb_est27.*
28: gb_est28.*
29: gb_est29.*
30: gb_est30.*
31: gb_est31.*
32: gb_est32.*
33: gb_est33.*
34: gb_est34.*
35: gb_est35.*
36: gb_est36.*
37: gb_est37.*
38: gb_est38.*
39: gb_est39.*
40: gb_est40.*
41: gb_est41.*
42: gb_est42.*
43: gb_est43.*
44: gb_est44.*
45: gb_est45.*
46: gb_est46.*
47: gb_est47.*

em_esthum10.*
45: em_esthum11.*
46: em_esthum12.*
47: em_esthum13.*
48: em_esthum14.*
49: em_esthum15.*
50: em_esthum16.*
51: em_esthum17.*
52: em_esthum18.*
53: em_esthum19.*
54: em_esthum20.*
55: em_esthum21.*
56: em_esthum22.*
57: em_esthum23.*
58: em_esthum24.*
59: em_esthum25.*
60: em_esthum26.*
61: em_esthum27.*
62: em_esthum28.*
63: em_estin1.*
64: em_estin2.*
65: em_estin3.*
66: em_estin4.*
67: em_estin5.*
68: em_estom1.*
69: em_estom2.*
70: em_estov1.*
71: em_estov2.*
72: em_estp11.*
73: em_estp12.*
74: em_estp13.*
75: em_estp14.*
76: em_estp15.*
77: em_estp16.*
78: em_estp17.*
79: em_estp18.*
80: em_estp19.*
81: em_estp110.*
82: em_estro1.*
83: em_estro2.*
84: em_estro3.*
85: em_estro4.*
86: em_estro5.*
87: em_estro6.*
88: em_estro7.*
89: em_estro8.*
90: em_estro9.*
91: em_estro10.*
92: em_estro11.*
93: em_estro12.*
94: em_estro13.*
95: em_estro14.*
96: em_estro15.*
97: em_estro16.*
98: em_estro17.*
99: em_estro18.*
100: em_estro19.*
101: em_estro20.*
102: gb_est25.*
103: gb_est26.*
104: gb_est27.*
105: gb_est28.*
106: gb_est29.*
107: gb_est30.*
108: gb_est31.*
109: gb_est32.*
110: gb_est33.*
111: gb_est34.*
112: gb_est35.*
113: gb_est36.*
114: gb_est37.*
115: gb_est38.*
116: gb_est39.*

117:	gb_est48:	*
118:	gb_est49:	*
119:	gb_est50:	*
120:	gb_est51:	*
121:	gb_est52:	*
122:	gb_est53:	*
123:	gb_est54:	*
124:	gb_est55:	*
125:	gb_est56:	*
126:	gb_est57:	*
127:	gb_est58:	*
128:	gb_est59:	*
129:	gb_est60:	*
130:	gb_est61:	*
131:	gb_est62:	*
132:	gb_est63:	*
133:	gb_est64:	*
134:	gb_est65:	*
135:	gb_est66:	*
136:	gb_est75:	*
137:	gb_est76:	*
138:	gb_est77:	*
139:	gb_est78:	*
140:	gb_est79:	*
141:	gb_est80:	*
142:	gb_est81:	*
143:	gb_est82:	*
144:	gb_est83:	*
145:	gb_est84:	*
146:	gb_est85:	*
147:	gb_est86:	*
148:	gb_est87:	*
149:	gb_est88:	*
150:	gb_est89:	*
151:	gb_est90:	*
152:	gb_est99:	*
153:	gb_est100:	*
154:	gb_est101:	*
155:	gb_est102:	*
156:	gb_est103:	*
157:	gb_est104:	*
158:	gb_est105:	*
159:	gb_est106:	*
160:	gb_est67:	*
161:	gb_est68:	*
162:	gb_est69:	*
163:	gb_est70:	*
164:	gb_est71:	*
165:	gb_est72:	*
166:	gb_est73:	*
167:	gb_est74:	*
168:	gb_est91:	*
169:	gb_est92:	*
170:	gb_est93:	*
171:	gb_est94:	*
172:	gb_est95:	*
173:	gb_est96:	*
174:	gb_est97:	*
175:	gb_est98:	*
176:	em_esthum29:	*
177:	em_esthum30:	*
178:	em_esthum31:	*
179:	em_esthum32:	*
180:	em_esthum33:	*
181:	em_estom3:	*
182:	em_estpl11:	*
183:	em_estro21:	*
184:	em_estro22:	*
185:	em_estro23:	*
186:	em_hlc:	*
187:	gb_est107:	*
188:	gb_est108:	*
189:	gb_est109:	*

190:	gb_estll0.*
191:	gb_estll1.*
192:	gb_hlc.*
193:	em_gss_fun.*
194:	em_gss_hum1.*
195:	em_gss_hum2.*
196:	em_gss_hum3.*
197:	em_gss_hum4.*
198:	em_gss_hum5.*
199:	em_gss_hum6.*
200:	em_gss_hum7.*
201:	em_gss_hum8.*
202:	em_gss_hum9.*
203:	em_gss_inv1.*
204:	em_gss_inv2.*
205:	em_gss_inv3.*
206:	em_gss_other.*
207:	em_gss_pln1.*
208:	em_gss_pln2.*
209:	em_gss_pro.*
210:	em_gss_rdd1.*
211:	em_gss_rdd2.*
212:	em_gss_rdd3.*
213:	em_gss_rdd4.*
214:	em_gss_rdd5.*
215:	em_gss_rdt1.*
216:	em_gss_vrt1.*
217:	em_gss_vrt3.*
218:	gb_gss1.*
219:	gb_gss2.*
220:	gb_gss3.*
221:	gb_gss4.*
222:	gb_gss5.*
223:	gb_gss6.*
224:	gb_gss7.*
225:	gb_gss8.*
226:	gb_gss9.*
227:	gb_gss10.*
228:	gb_gss11.*
229:	gb_gss12.*
230:	gb_gss13.*
231:	gb_gss14.*
232:	gb_gss15.*
233:	gb_gss16.*
234:	gb_gss17.*
235:	gb_gss18.*
236:	gb_gss19.*
237:	gb_gss20.*
238:	gb_gss21.*
239:	gb_gss22.*
240:	gb_gss23.*
241:	gb_gss24.*
242:	gb_gss25.*
243:	gb_gss26.*
244:	gb_gss27.*
245:	gb_gss28.*
246:	gb_gss29.*
247:	gb_gss30.*
248:	gb_gss31.*
249:	gb_gss32.*
250:	gb_gss33.*
251:	gb_gss34.*
252:	em_gss_inv4.*
253:	em_gss_rdd6.*
254:	em_gss_rdd7.*
255:	gb_gss35.*
256:	gb_gss36.*
257:	gb_gss37.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Db	90	CAGCGACGCGCGTGGTGGCGGAGCAGGACCGCGGTGATGCGCGATCTCGAAGCATGTCGCGN	149
QY	216	atcatcgcgcggtgctgcgcgcacgcgaagatctcgagcagcgccaaaggagagcagtc	275
Db	150	ATCATGCGCGGTGCGTCTCCCTGCCACGCAAGATCTCCGACGACGACGCAANGAGCGATC	209
QY	276	caggagtgcgtgtcggagtagcatcagcttcatcacgggggagggccaaacagcagcgtgcag	335
Db	210	CAGGAGTGCGTGCGAGTTCATCAGCTTCGTCACCGCGAGGCAACTAGCGGTGCCAC	269
QY	336	cgggagcagcgaagacaatcaccccgagagcgtgctgtgggccaatgagccctcctgcgc	395
Db	270	ATGAGCAGCCGCAAGACCGCTCAACCGGAAGACATCTGTTGGGCCCTGAACACGATCGC	329
QY	396	ttcagcagctactgcagcgcctcgccgacctacctccacgcgtaccgcaggttcgagggc	455
Db	330	TTGCGAGCTACTGCTCTGCCCTTCANCGTCTTCTGTCACCGATGCGCGAGGGGAGCGG	389
QY	456	gacgcgcgcgcgtcggtcgctcccgggggcgcccccacatgcgcg	501
Db	390	GGACAGGTGGTGGCGGTGTCAGGCGCAACACCGCGCGCCACGAGGG	435
RESULT	5		
BG551755			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
COMMENT			
FEATURES			
source			

Df 291 CTGAGTACATCAGCTTCATCACAGGTGAGCGCAACGAGCGTTGCCAGAGGAGCAQCGGA 350

Qy 349 agaccatcaccgagacgtactgttggccatgagcgctcgcttcgcagtacg 408
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Df 351 AGACCATTACCGAGAGAGACTGCTTTGGCCCATGAGCAAGCTTGATTCGACGACTACA 410
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 409 tgagcgctggcgcttaacctcaacctaccgagttccagggttgaggggagcgcg 462
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Df 411 TCGAACCGTTGACCATGTACCTTCACCGCTACCGTGAACCTTGAGGTGACCGCAC 464
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 9
AV424305/c 418 bp mRNA EST 23-MAY-2000
LOCUS AV424305 Lotus japonicus young plants (two-week old) Lotus
DEFINITION japonicus cDNA clone MWM038e02_r 5'', mRNA sequence.
ACCESSION AV424305
VERSION AV424305.1 GI:7781090
KEYWORDS EST.
SOURCE Lotus japonicus.
ORGANISM Lotus japonicus
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Lotus.
1 (bases 1 to 418)
REFERENCE Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
AUTHORS Generation of 7137 non-redundant expressed sequence tags from a
TITLE Legume, Lotus japonicus
JOURNAL DNA Res. 7 (2), 127-130 (2000)
MEDLINE 20277479
COMMENT Contact: Yasukazu Nakamura
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: ynakamu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
Location/Qualifiers
FEATURES
source
1..418
/organism="Lotus japonicus"
/db_xref="taxon:34305"
/clone="MWM038e02_r"
/clone_lib="Lotus japonicus young plants (two-week old)"
/dev_stage="young plants (two-week old)"
/note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
XhoI; isolate=Miyakojima MG-20"
BASE COUNT 63 a 155 c 102 g 98 t
ORIGIN

Query Match 16.1%; Score 189.2; DB 30; Length 418;
Best Local Similarity 68.8%; Pred. No. 6.5e-32;
Matches 260; Conservative 0; Mismatches 118; Indels 0; Gaps 0

Qy 118 cgcgcgcgcgccaatggcgcgctgctcagcagcatgcggcgccgagctccg 177
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Df 379 CCGGGGGAGACACACCGAGCGGATCGAAGCTCATGCAGAGAGTTATCGCTGAAGG 320
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 178 agcaggaccggctgatgcogattcggaacgttgtatccgcatactgcgcgctgcgcg 237
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Df 319 AGCAGGACCGGTTCTGCCGATAGCGAAAGCTGAGGAGGATCATGAAGAAGCGGTGCGCG 260
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 238 gcacgcagaatctgcagcagcgccaaaggagcacgtccaggagtgcgtgctcggaagtaca 297
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Df 259 CGAAGCGGAAGATCTCGAAGACCCGAGGACACGGTGCAGGAGTGCCTCGGAGTTCA 200
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 298 tcagcttcacgcggggagggccaaacgagcgtgtccagcggagcagcgaagaccatca 357
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Df 199 TAAGCTTCATCACC GGCGAGGCGCTCCGACAAGTGGCCAGCGGGAAGGAGACGATCA 140
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 358 ccgcgcagagacgtactgtggccatatagcgcctcggttgcagcactacgtcgagcgc 417
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Df 139 ACGCGGATGACCTGCTGTGGGCCATGACGACGGTTGGGTTTCGAGGACTACGTGAGCGCG 80
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Ov 418 tcggcgctactctccacgcgtacccgagttcgagggcgagcgcgcggtcggtcgctcg 477


```

/organism="Glycine max"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-c1027-1821"
/clone_lib="Gm-c1027"
/tissue_type="cotyledons of 3- and 7-day-old Williams
seedlings"
/lab_host="DH10B"
/Note="Vector: pBluescript II SK4; Site_1: EcoRI; Site_2:
XhoI; This cDNA library was constructed from mRNA isolated
from cotyledons of 3- and 7-day-old Williams seedlings
which were propagated on paper towels with distilled
water. The cotyledons were flash-frozen in liquid
nitrogen, then lyophilized for 72 hours. Unequal amounts
of mRNA was used for cDNA synthesis. Stratagene's cDNA
Synthesis kit (catalog number 200401) was used to
synthesize the cDNA. First- strand synthesis was
performed with 5-methyl dCp, hence the ligated cDNA was
hemimethylated. A modification of Stratagene's
first-strand synthesis primer was used. An anchor
nucleotide (V-A, C, or G) was added to the 3' end of the
primer [GAGAGAGAGAGAGAGACTACTGCGAT(18)] to anchor
the primer at the 5' end of the poly(A) tract. After
second- strand synthesis, the cDNA ends were filled in

```


OM of: US-09-435-054-2 to: EST:* out_format : pfs

Date: Oct 25, 2001 2:03 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=frame+2n.model -DEV=xlh
-Q/cg2_1/USP02.spool/US09435054/runat_25102001_085522.6964/app_query.fasta_1.338
-DB=EST -QFMT=fastap -SUFFIX=p2n.rst -GAPOP=12.000 -CAPEXT=4.000
-MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500
-QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
-FGAPEXT=7.000 -START=1 -XGAPEXT=0.500 -DELOP=6.000
-DELEXT=7.000 -START=1 -XGAPEXT=0.500 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0
-MAXLEN=200000000 -USER=US09435054.ecgN1.1.3889 -NCPU=6
-ICPU=3 -LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY
-WAIT_THREADS=1

Search information block:

Query: US-09-435-054-2
Query length: 278
Database: EST.*
Database sequences: 10228115
Database length: 431459454
Search time (sec): 1006.350000

score_list:

Sequence	Strd	Orig	Zscore	EScore	Len	Documentation
gb_est30:AU088581	+	771.00	1053.30	1.5e-49	695	! AU088581 AU088581 Rice callus C
gb_est30:BE603222	+	598.50	819.72	1.5e-36	528	! BE603222 HVSMH0102116f Hordeum
gb_est51:AW756413	+	481.00	657.74	1.6e-27	591	! AW756413 s12a12.y1 Gm-cl036 G
gb_est102:BG551755	+	465.50	638.18	2.0e-26	497	! BG551755 s242f11.y1 Gm-cl075 G
gb_est47:AW432980	+	453.50	637.19	2.3e-26	551	! AW432980 si03a01.y1 Gm-cl029 G
gb_est30:BF593304	+	465.50	622.29	1.5e-25	470	! BF593304 su76f03.y1 Gm-cl055 G
gb_est28:AL506199	+	446.50	610.45	7.0e-25	594	! AL506199 AL506199 Hordeum vulg
gb_est51:AW754604	+	431.00	590.23	9.3e-24	535	! AW754604 PC04B12 Pine Triplex F
gb_est54:AW981720	+	422.00	578.07	4.4e-23	526	! AW981720 PC15H07 Pine Triplex F
gb_est38:AL509098	+	417.50	573.60	7.9e-23	441	! AL509098 AL509098 Hordeum vulg
gb_est74:BF418716	+	410.00	556.23	7.3e-22	924	! BF418716 SC10A74.B01R990724 IPE
gb_est27:AL1995140	+	406.50	555.99	7.5e-22	575	! AL1995140 701502287 A. thaliana
gb_est85:BF263455	+	405.00	553.61	1.0e-21	595	! BF263455 HV_CEA0006M10f Hordeum
gb_est95:BF263455	+	403.00	553.61	1.0e-21	595	! BF263455 HV_CEA0006M10f Hordeum
gb_est39:AV550943	+	403.00	551.43	1.3e-21	561	! AV550943 HV_CEA0006M16f Hordeum
gb_est38:AV424305	+	401.00	551.52	1.3e-21	418	! AV424305 AV424305 Arabidopsis t
gb_est79:BE803572	+	400.00	549.05	1.8e-21	469	! BE803572 sr60e11.y1 Gm-cl052 G
gb_est100:BG440251	+	400.00	545.40	2.9e-21	686	! BG440251 GA_Ea0006K20f Gossyp
em_estp15:AW666992	+	400.00	545.29	3.0e-21	694	! AW666992 GA_Ea0006K20 Gossyp
gb_est69:BE021941	+	397.00	543.80	3.6e-21	528	! BE021941 sm64d05.y1 Gm-cl028 G
gb_est24:AL1725612	+	394.00	537.48	8.1e-21	665	! AL1725612 BNLGH112445 Six-day C
gb_est99:BG363233	+	393.50	538.76	6.9e-21	542	! BG363233 sac1111.y1 Gm-cl040 G
gb_est46:AW395227	+	393.50	538.02	7.5e-21	585	! AW395227 sh45e04.y1 Gm-cl017 G
gb_est91:BF715909	+	393.00	540.27	5.6e-21	431	! BF715909 saalle08.y1 Gm-cl058 G
gb_est50:AW719547	+	392.00	537.71	7.8e-21	488	! AW719547 LjNEST6a3r Lotus japon
gb_est50:AW720671	+	392.00	536.90	8.7e-21	531	! AW720671 LjNEST6a3r Lotus japon
gb_est49:AW597630	+	392.00	536.30	9.4e-21	565	! AW597630 sj96q06.y1 Gm-cl023 G
gb_est47:AW459387	+	391.50	537.49	8.1e-21	465	! AW459387 sh23f03.y1 Gm-cl016 G
gb_est100:BG445358	+	391.50	531.49	1.7e-20	869	! BG445358 GA_Ea00027N18f Gossyp
em_estp15:AW730639	+	391.50	531.40	1.8e-20	877	! AW730639 GA_Ea00027N18f Gossyp
gb_est99:BG314203	+	390.50	533.08	1.4e-20	638	! BG314203 WHE2460.E10.T202s Trit
gb_est74:BE414135	+	390.00	532.92	1.4e-20	604	! BE414135 WHE2460.E10.T202s Trit
gb_est71:BE210041	+	388.50	533.61	1.3e-20	454	! BE210041 so38b01.y1 Gm-cl039 G
gb_est90:BF597252	+	388.50	530.68	1.9e-20	616	! BF597252 su96c06.y1 Gm-cl056 G
gb_est75:BE516510	+	387.50	529.68	2.2e-20	593	! BE516510 WHE611.D10.H192a Wheat
gb_est51:AW75623	+	387.00	530.40	2.0e-20	512	! AW75623 SCU003.4688 DSTL Medicag
gb_est74:BE413647	+	387.00	527.13	3.0e-20	720	! BE413647 SCU003.E10.R990714 IPE
gb_est51:AW733618	+	386.50	529.17	2.3e-20	542	! AW733618 sk75h06.y1 Gm-cl016 G
gb_est21:AW1495007	+	385.50	529.61	2.2e-20	449	! AW1495007 sa89f03.y1 Gm-cl004 G
gb_est69:BE060015	+	384.00	529.57	2.2e-20	364	! BE060015 sn39h06.y1 Gm-cl027 G

gb_est26:AI900024 + 384.00 524.26 4.4e-20 633 ! AI900024 sb97q11.y1 Gm-cl012
gb_est54:AW980494 + 384.00 523.54 4.8e-20 682 ! AW980494 EST391647 GVN Medic
gb_est51:AW738727 + 383.50 523.91 4.6e-20 611 ! AW738727 EST340154 tomato fl
gb_est51:AW760103 + 383.00 526.78 3.2e-20 422 ! AW760103 sl58b03.y1 Gm-cl027
gb_est49:AW621652 + 382.50 523.89 4.6e-20 531 ! AW621652 EST312450 tomato ro

seq_name: gb_est30:AU088581

seq_documentation_block:
LOCUS AU088581 695 bp mRNA EST 31-MAR-2000
DEFINITION AU088581 Rice callus Oryza sativa subsp. japonica cDNA clone C52742
, mRNA sequence.

ACCESSION AU088581
VERSION AU088581.1 GI:7378310
KEYWORDS EST.
SOURCE Oryza sativa subsp. japonica.
ORGANISM Oryza sativa subsp. japonica

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.
1 (bases 1 to 695)
Sasaki,T. and Yamamoto,K.
Rice cDNA from callus (2000)
Unpublished (2000)

CONTACT: Takuji Sasaki
National Institute of Agrobiological Resources
Rice Genome Research Program
2-1-2 Kannondai,Tsukuba
Ibaraki,
Japan 305

Tel: 0298-38-7441
Fax: 0298-38-7468
Email: tsasaki@abr.affrc.go.jp
PROJECT -"RGP".
C52742_12A.

FEATURES
source
1..695
/organism="Oryza sativa subsp. japonica"
/strain="cultivar Nipponbare, sub-species Japonica"
/db_xref="taxon:39947"
/clone="C52742"
/note="Vector: pBluescript II SK+; Site_1: SalI; Site_2:
NotI; cDNA prepared from rice callus mRNAs by using
oligo(dT) as a primer and ligating to the SalI-NotI site
of pBluescript II SK+ phagemid."
BASE COUNT 124 a 245 c 232 g 88 t 6 others
ORIGIN

alignment_scores:
Quality: 771.00 Length: 239
Ratio: 4.016 Gaps: 10
Percent Similarity: 80.335 Percent Identity: 70.293

alignment_block:
US-09-435-054-2 x AU088581 ..
Align seg 1/1 to: AU088581 from: 1 to: 695

33 ProAlaileArgGluGlnAspArgLeuMetProileAlaAsnValilear 49
|||||
2 CCGCCTATACGTGAGCAGACCGCGTGTGCGGATCGGACGTATCGG 51
49 gileMetArgGluValLeuProAlaHisAlaLysleSerAspAlaL 66
|||||
52 CATCATGCGCGCGTCTCTCCGCGCAGCCAGATCTCGGACGCGCA 101
66 ysgLthrIleGlnGluCysValSerGluTyrIleSerPheIleThrGly 82
|||||
102 AGGAGACGATCAGGAGTCCGTGTCGAGATACATCATCATCCGCG 151
83 GluAlaAsnGluArgCysGlnArgGluGlnArgLysThrIleThrAlaGl 99

```

|||||
152 GAGGCCAACGAGCGGTGCCAGCGAGCGAGCAAGACCATCACCCCGCA 201
99 uaspValLeuTrpAlaMetSerArgLeuGlyPheAspAspTyrValGlup 116
|||||
202 GGAGCTGCTCTGGGCCATGAGCGCGCTCGGCTTCGAGCACTAGTCGAGC 251
116 roLeuGlyAlaTyrLeuHisArgTyrArgGluPheGluGlyAspAlaArg 132
|||||
252 CCCTCGCGCGTCTACCTCCACCGCTACCGGAGTTCGAGGGGAGTCCCGC 301
133 GlyValGlyLeuValProGlyAlaAlaProSerArgGlyGlyAspHisHi 149
|||||
302 GGCCTCGCGCTCGCGCTCGCGCGCGG.....CGCGGCGACACCA 345
149 sProHisSerMetSerProAlaAlaMetLeuLysSerArgGlyProValS 166
|||||
346 TGGTCAC.....GTCCGTGGGATGCTCAAGTCCCGCGCGCAGGGCT 386
166 erGlyAlaAlaMetLeuProHisHisHisHisHisHisMetGlnMet 182
|||||
387 CCATGGTGAGC.....CACCACGACATGCAGATG 415
183 HisAlaAlaMetTyrGlyGlyThrAlaValProProProAlaGlyProPr 199
|||||
416 CACGCCCCCATGTACGGTGGCGCGCGGTGCCCGCGCGCATCTCC 465
199 o..HisHisGlyGlyPhe.....LeuMetProHisProGlnGlySer 213
|||||
466 TCCGCCACACCGGTTCCACCGCTCATGCCG.....GC 500
213 rHisTyrLeuProTyrAla.....TyrGluProThrTyrGlyGlyLuh 228
|||||
501 GCACACCGGCGAGTACGCGCGCGCGTACGAC...ATGTACGCGCGGAGC 547
228 isAlaMetAlaAlaTyrTyrGlyGlyAlaAlaProGlyAsnGly 244
|||||
548 ACGGGATGGCGGGCTACTACGGCGG...ATGTACGCGCGCGCGCGCGC 594
245 GlySerGlyAspGlySerGlySerGlyGlyGlyGly.....Se 258
|||||
595 GGCNACGGGAGCGGACGCGCGCGGTGGCGCGGACCGCGNAGAA 644
258 rAlaSerHisThrPro 263
|||||
645 CGTCAATTNACACA 660

```

seq_name: gb_est76:BE603222

```

seq_documentation_block: 528 bp mRNA EST 02-MAR-2001
LOCUS BE603222
DEFINITION HVSMeh0102J16f Hordeum vulgare 5-45 DAP spike EST library
HVCNDA0009 (5 to 45 DAP) Hordeum vulgare cDNA clone HVSMeh0102J16f,
mRNA sequence.
ACCESSION BE603222
VERSION BE603222.2 GI:13191083
KEYWORDS EST.
SOURCE barley.
ORGANISM Hordeum vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
; Triticeae; Hordeum.
1 (bases 1 to 528)
Wing,R., Close,T.J., Kleinhofs,A., Wise,R., Begum,D., Frisch,D., Yu
,Y., Anderson,H., Dale,J., Henry,D., Kernodle,S., Palmer,M., Rambo
,T., Saski,C., Schwartzbeck,J., Simmons,J., Choi,D.W., Main,D. and
Wood,T.
Development of a genetically and physically anchored EST resource
for barley genomics
Unpublished (2000)
JOURNAL On Aug 21, 2000 this sequence version replaced gi:9860783.
COMMENT Contact: Wing RA
Clemson University Genomics Institute

```

```

Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Seq primer: AATTACCCCTCACTAAAGG
High quality sequence stop: 512.
Location/Qualifiers
source
1..528
/organism="Hordeum vulgare"
/cultivar="Morex"
/db_xref="taxon:4513"
/clone="HVSMeh0102J16f"
/clone_lib="Hordeum vulgare 5-45 DAP spike EST library"
/HVCNDA0009 (5 to 45 DAP)"
/tissue_type="5-45 DAP Spike"
/lab_host="SOLR"
/note="Vector: lambdaZAP; Site_1: EcoRI; Site_2: XhoI"

BASE COUNT 97 a 181 c 182 g 68 t
ORIGIN

alignment_scores:
Quality: 598.50 Length: 174
Ratio: 4.215 Gaps: 5
Percent Similarity: 81.609 Percent Identity: 71.264

alignment_block:
US-09-435-054-2 x BE603222 ..
Align seg 1/1 to: BE603222 from: 1 to: 528

```

```

11 GlyAlaGluAsnGlySerAlaAlaGlyGlyAlaAsnAsnGlyGlyAlaAl 27
|||||
62 GCGCGCGCGAAGCGCACGTCTTCGCGCGCGCGGGAACGGTGGCGCGG 111
27 agGlnHisAlaAlaProAlaAlaArgGluGlnAspArgLeuMetProI 44
: : : : : : : : : : : : : : : : : : : : : : : : : : :
112 CCGCAT.....GCGCGCGCGTTCGCGAGCAGGAGGTGTGATGCCGA 155
44 leAlaAsnValIleArgIleMetArgValLeuProAlaHisAlaLys 60
|||||
156 TCGGACGTCGTCGTCATCATCGCGCGCGTCTCGCGCGCGCGCGCAAG 205
61 leSerAspAlaLysGluThrIleGlnGluCysValSerGluTyrI 77
|||||
206 ATCTCGGACGCGCAAGGAGAGCATCCAGGAGTGGTGTCTCGAGTACAT 255
77 eSerPheIleThrGlyGluAlaAsnGluArgCysGlnArgGluGlnArgL 94
|||||
256 CAGCTTCATCAGCGGGAGGCGCAATGAGCGGTGCCAGCGCGAGCGCA 305
94 ysThrIleThrAlaGluAspValLeuTrpAlaMetSerArgLeuGlyPhe 110
|||||
306 AGACCATCACCGCGGAGCGTCTCTGGCGCATGAGCGCGCTCGGCTTC 355
111 AspAspTyrValGluProLeuGlyAlaTyrLeuHisArgTyrArgGluPh 127
|||||
356 GACGACTAGTGTGAGCGCGCTCAGCATCTACCTCCACCGCTTCGCGAGTT 405
127 eGluGlyAspAlaArgGlyValGlyLeuValProGlyAlaAlaProSerA 144
406 CAGGGCGGAGGCGCGCGCGCGCGT..... 433
144 rgGlyGlyAspHisHisProHisSerMetSerProAlaAlaMetLeuLys 160
|||||
434 .....GGCCACCATCACCGGAGT.....CGCCCATGATGCTCAAG 472
161 SerArgGlyProValSerGlyAlaAlaMetLeuProHisHisHisHis 177
|||||
473 CCGCGCGG.....GCGCGCGGCGCATGGTGGCG.....CA 504
177 shisAspMetGlnMetHisAla 184

```

Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Seq primer: AATTACCCCTCACTAAAGG
High quality sequence stop: 512.
Location/Qualifiers
source
1..528
/organism="Hordeum vulgare"
/cultivar="Morex"
/db_xref="taxon:4513"
/clone="HVSMeh0102J16f"
/clone_lib="Hordeum vulgare 5-45 DAP spike EST library"
/HVCNDA0009 (5 to 45 DAP)"
/tissue_type="5-45 DAP Spike"
/lab_host="SOLR"
/note="Vector: lambdaZAP; Site_1: EcoRI; Site_2: XhoI"

COMMENT

Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available through: Genome Systems, Inc. 4633 World
Parkway Circle St. Louis, Missouri 63134 For further information
call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314)
427-3324 or contact: clones@genomesystems.com or
info@genomesystems.com web site: www.genomesystems.com
High quality sequence stop: 370.

FEATURES

Location/Qualifiers

1..497

/organism="Glycine max"

/db_xref="taxon:3847"

/clone="GENOME SYSTEMS CLONE ID: Gm-cl075-669"

/clone_lib="Gm-cl075"

/tissue_type="differentiating somatic embryos cultured on MSM6AC"

/lab_host="DH10B"

/note="Vector: pBluescript II SK+; Site_1: EcoRI; Site_2: XhoI; The cDNA library was constructed from mRNA isolated from differentiating somatic embryos cultured on MSM6AC. The library was prepared using the Stratagene pBluescript II SK(+) library construction kit. Complementary DNA was synthesized from mRNA using a primer consisting of a poly(dT) sequence with an XhoI restriction site. EcoRI adaptors were ligated to the blunt-ended cDNA fragments followed by XhoI digestion. The cDNA fragments were directionally cloned into the EcoRI-XhoI restriction site of the pBluescript vector. The ligated cDNA fragments were transformed into E.coli Electromax DH10B host cells. Tissue culture and library construction were performed by Françoise Thibaud-Nissen and Anu Khana (Lila Vodkin lab, University of Illinois)."

BASE COUNT 146 a 130 c 131 g 90 t

ORIGIN

alignment_scores

Quality: 465.50 Length: 121

Ratio: 4.392 Gaps: 1

Percent Similarity: 87.603 Percent Identity: 74.380

alignment_block

US-09-435-054-2 x BG551755 ..

Align seg 1/1 to: BG551755 from: 1 to: 497

10 AlaGlyAlaGluAsnGlySerAlaAlaGlyGlyAlaAsnAsnGlyGlyAl 26

138 GCATCATCATGATCAGTCAGTCAGCCACAGGAGGAGACGAA..... 179

26 aAlaGlnGlnHisAlaAlaProAlaIleArgGluGlnAspArgLeuMetp 43

180TGCACGGTCAGGAGCAAGCAGAGGTTTCATGC 210

43 rolleAlaAsnValIleArgIleMetArgGValValLeuProAlaHisAla 59

211 CAATCGCTAACGTGATTAGGATCATGCGCAAGATTCTCCCTCCACACCCA 260

60 LysIleSerAspAlaLysGluThrIleGlnGluCysValSerGluTyr 76

261 AAAATCTCGAGCATGCAAAAGCAAACTCAAGAGTCGTCTGTGAGTA 310

76 rIleSerPheIleThrGlyGluAlaAsnGluArgCysGlnArgGluGlnA 93

311 CATCATGCTTCATCATCAGGTGAGGCGCAAGCAGGCTTGCACAGGGAGCAGC 360

93 rGlySerThrIleThrAlaGluAspValLeuTyrAlaMetSerArgLeuGly 109

110 PheAspAspTyrValGluProLeuGlyAlaTyrLeuHisArgTyrArgG1 126

411 TTCGACGACTACATGACACCGTTTGACCATGTACCTTCACCGCTACCGTGA 460

126 uPheGluGlyAsp 130

461 ACTTGAGGCGTGAC 473

seq_name: gb_est47:AW432980

seq_documentation_block:

LOCUS AW432980 551 bp mRNA 18-JUL-2000

DEFINITION si03a01.y1 Gm-cl029 Glycine max cDNA clone GENOME SYSTEMS CLONE ID: Gm-cl029-97 5' similar to TR:081130 081130 CCAAT-BOX BINDING FACTOR HAP3 HOMOLOG. ; mRNA sequence.

ACCESSION AW432980

VERSION AW432980.1 GI:6964287

KEYWORDS EST.

SOURCE soybean.

ORGANISM Glycine max

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Glycine. 1 (bases 1 to 551)

AUTHORS Shoemaker,R., Keim,P., Vodkin,L., Erpelting,J., Coryell,V., Khanna,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Wilson,R.

PUBLIC Soybean EST Project

Public Soybean EST Project

Contact: Shoemaker R/Public Soybean EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

This clone is available through: Genome Systems, Inc. 4633 World Parkway Circle St. Louis, Missouri 63134 For further information call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314) 427-3324 or contact: clones@genomesystems.com or info@genomesystems.com web site: www.genomesystems.com

Insert Length: 1007 Std Error: 0.00

High quality sequence stop: 455.

Location/Qualifiers

1..551

/organism="Glycine max"

/db_xref="taxon:3847"

/clone="GENOME SYSTEMS CLONE ID: Gm-cl029-97"

/clone_lib="Gm-cl029"

/tissue_type="very young cotyledons of greenhouse grown plants"

/lab_host="DH10B"

/note="Vector: pSPORT1; Site_1: Not I; Site_2: Sal I; This cDNA library was constructed from mRNA isolated from very young cotyledons (20-50mg fresh weight) of greenhouse grown plants. The library was prepared using the Life Technologies pSuperScript cDNA library construction kit. Complementary DNA was synthesized from mRNA using a poly(dT) sequence with a Not I restriction site. Sal I linker adaptors were ligated to the blunt-ended cDNA fragments followed by Not I digestion. The cDNA fragments were directionally cloned into the Not I-Sal I restriction site of the pSPORT1 vector. The ligated cDNA fragments were transformed into E.coli Electromax DH10B host cells. This library was constructed by Dr. Paul Keim and Dr. Virginia Coryell."

BASE COUNT 156 a 147 c 144 g 104 t

ORIGIN

448 ACTTGAG 454

seq_name: gb_est28.AL506199

seq_documentation_block:

LOCUS AL506199 594 bp mRNA EST 04-JAN-2001
 DEFINITION Hordeum vulgare Barke developing caryopsis (3'-15.DAP)
 Hordeum vulgare cDNA clone HY02F18T 5', mRNA sequence.

ACCESSION AL506199.1 GI:12032414

VERSION AL506199.1

KEYWORDS EST.

SOURCE barley.

ORGANISM Hordeum vulgare

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae

1 (bases 1 to 594)

AUTHORS Michalek, W., Weschke, W., Pleissner, K.-P. and Graner, A.

TITLE EST sequencing and analysis in barley

JOURNAL Unpublished (2000)

COMMENT Contact: Michalek W

Institute for Plant Genetics and Crop Plant Research

Corrensstr.3, D-06466 Gatersleben, Germany

Email: michalek@ipk-gatersleben.de, http://pgrc.ipk-gatersleben.de

Seq primer: T3 primer for 5' end.

FEATURES

1..594
 source

/organism="Hordeum vulgare"

/cultivar="Barke"

/db_xref="taxon:4513"

/clone_lib="Hordeum vulgare Barke developing caryopsis

(3'-15.DAP)"

/tissue_type="developing caryopsis (3'-15.DAP)"

/lab_host="XL0UR"

/note="Vector: plasmid pBK-CMV; Site_1: EcoRI; Site_2:

XhoI; mRNA was made from developing caryopsis (3'-15.DAP)

of spring barley variety 'Barke', a high quality malting

variety. Cloning sites: EcoRI (5'-end of cDNA) and XhoI

(3'-end of cDNA). NOTE: Due to a cloning artefact caused

by the kit, in most cases the EcoRI site is NOT present,

as well as the EcoRI adapter. Average insert size is 1 kb

Sequence trimming: Vector sequences and sequence ends were

trimmed from the 5'- and 3'-end until a 50 bp window

contains less than two ambiguities. The maximum length was

set to 700 bp"

BASE COUNT 109 a 213 c 185 g 85 t 2 others

ORIGIN

alignment_scores:

Quality: 446.50 Length: 193

Ratio: 3.212 Gaps: 4

Percent Similarity: 72.021 Percent Identity: 53.368

alignment_block:

US-09-435-054-2 x AL506199 ..

Align seg 1/1 to: AL506199 from: 1 to: 594

9 AlaAlaGlyAlaGlnGlySerAlaAlaGlyAlaAsnAspGlyG1 25

58 GCCCTAGCCATGGACGATTCACCGTCCGACGAGCA.....CGCGC 101

25 yAlaAlaGlnGlnHisAlaAlaProAlaAlaArgGluGlnAspArgLeuM 42

102 GCCGGGCCCACTCAGCGACCGCGTGTGGGAGCAGGACCGCGTGA 151

42 etProTleAlaAsnValIleArgIleMetArgValValLeuProAlaHis 58

152 TGCCGATCGGAACGATTCACCGATCATCGCGTCCCTCCCTCCCGCAC 201

59 AlaLysIleSerAspAspAlaLysGluThrIleGlnGluCysValSerG1 75

|||||

202 GCCAAGATCTCCGACGACGCCAAGGAGCGATCCAGGAGTGGTGTCCGA 251

75 wTyTleSerPheIleThrGlyGluAlaAsnGluArgCysGlnArgGluG 92

252 GTTCATCAGCTTCGTACCGGAGGCAACAGGCGGTGCCACATGGAGC 301

92 lnArgLysThrIleThrAlaGluAspValLeuTrpAlaMetSerArgLeu 108

302 ACCGCAAGACCGTCAACGCGGAAGACATCGTGTGGCGCTGAACCGCTC 351

109 GlyPheAspAspTyrValGluProLeuGlyAlaTyrLeuHisArgTyrAr 125

352 GCGTTCGACGACTAGCTCGTCCCTCAGCGCTTCCTGCGACCGCATCGG 401

125 gGluPheGluGlyAspAlaArgGlyValGlyLeuValProGlyAlaAla 142

402 CAGAGGCGAGGCGGACAGGTGGTGGCGGTGCAGCGACACACCGCGCG 451

142 roSerArgGlyGlyAspHisHisProHisSerMetSerPro..... 155

452 CCAGGAGGCGCGCTCCCGCGCGCGGCC.GNCCGCTCTCCACGCGGTGCC 500

156AlaAlaMetLeuLysSerArgGlyProValSerGlyAlaAl 169

501 GTTGCAGGTACTGCAAGCGTCCATGTACGCGCCCGCTCTCCGTTGCAGC 550

169 aMetLeuProHisHisHisHisAspMetGlnMetHisAlaAlaAlam 186

551 T.....CAGAATCACATTGCNA 567

186 etTyrGlyGlyThrAlaValProProPro 195

568 GA.....GGCCAGTTGTACGCGCGCCCCC 590

seq_name: gb_est51:AW754604

seq_documentation_block:

LOCUS AW754604 535 bp mRNA EST 01-MAY-2000

DEFINITION PC04B12 Pine Triplex pollen cone library Pinus taeda cDNA clone

PC04B12, mRNA sequence.

ACCESSION AW754604

KEYWORDS EST.

SOURCE loblolly pine.

ORGANISM Pinus taeda

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus; Pinus.

Whetten, R.W., Kinlaw, C.S., Retzel, E. and Sederoff, R.R.

The Pine Gene Discovery Project

Unpublished (1999)

Contact: Ross Whetten

Forest Biotechnology Group

North Carolina State University

Dept. of Forestry, NC State University, 6113 Jordan Hall, Raleigh

NC, 27695-8008

Tel: 919-515-7800

Fax: 919-515-7801

Email: rosswhet@unity.ncsu.edu

Seq primer: 5' lambda triplex2 Sequencing Primer.

Location/Qualifiers

1..535

/organism="Pinus taeda"

/db_xref="taxon:3352"

/clone_lib="Pine Triplex pollen cone library"

/dev_stage="immature"

/lab_host="E. coli BM25.8"

/note="Organ: pollen cone; Vector: Lambda Triplex; Site_1:

SfiI (A); Site_2: SfiI (B); Immature pollen cones were

collected in the early spring, frozen and used for mRNA

isolation. The SMART-PCR method (Clontech) was used to

prepare a library from 1 ug total RNA, using the Lambda
Triplex vector. Plasmid subclones in pTriplex were
recovered by cre-lox excision in E. coli strain BM25.8 and
sequenced from the 5' end."

BASE COUNT 143 a 103 c 129 g 139 t 21 others
ORIGIN

alignment_scores:
Quality: 431.00 Length: 99
Ratio: 4.585 Gaps: 0
Percent Similarity: 94.949 Percent Identity: 82.828

alignment_block:
US-09-435-054-2 x AW754604 ..

Align seg 1/1 to: AW754604 from: 1 to: 535

```

34 AlaIleArgGluGlnAspArgLeuMetProIleAlaAsnValIleArgII 50
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
188 GCTGTGAGAGACAAGATAGTGTTCATGCCCATTCCTAATGTCATTAGGAT 237
50 eMetArgArgValLeuProAlaHisAlaLysIleSerAspAspAlaLysG 67
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
238 AATGAGAAAGTTCTACCCACCCATGCAAGATTTCTGATGATGCAAGG 287
67 luThrIleGlnGluCysValSerGluTyrIleSerPheIleThrGlyGlu 83
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
288 AGACCATTTCAAGAAATGTGTCTGATGATACATAAGCTTCATCACCAGTGA 337
84 AlaAsnGluArgCysGlnArgGluGlnArgLysThrIleThrAlaGluAs 100
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
338 GCGACAGATGCTGCGCAAGAGGAGCAGAGAAAACACTATCATCTGCTGA 387
100 pValLeuTrpAlaMetSerArgLeuGlyPheAspTyrValGluProL 117
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
388 TGCTTCTCGGCTATGACCAAGTTGGGTTTGTGATGATATGATGAGGCAT 437
117 euGlyAlaTyrIleuHisArgTyrArgGluPheGluGlyAspAlaArg 132
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
438 TAACTATTTTACCTCCAGAAATACAGAGATGCTGAAGNNNNNNNCA 484

```

seq_name: gb_est54:AW981720

seq_documentation_block: 526 bp mRNA EST 02-JUN-2000
LOCUS AW981720
DEFINITION PC15H07 Pine triplex pollen cone library Pinus taeda cDNA clone
PC15H07, mRNA sequence.
ACCESSION AW981720
VERSION AW981720.1 GI:8173288
KEYWORDS EST.
SOURCE loblolly pine.
ORGANISM Pinus taeda

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Coniferales; Pinaceae; Pinus; Pinus.
1 (bases 1 to 526)
Whetten, R.W., Kinlaw, C.S., Retzel, E. and Sederoff, R.R.
The Pine Gene Discovery Project
Contact: Ross Whetten
Forest Biotechnology Group
North Carolina State University
Dept. of Forestry, NC State University, 6113 Jordan Hall, Raleigh
, NC, 27695-8008
Tel: 919-515-7800
Fax: 919-515-7801
Email: rosswhet@unity.ncsu.edu
Seq primer: 5' lambda Triplex2 Sequencing Primer.

FEATURES
source
1..526
Location/Qualifiers
/organism="Pinus taeda"
/db_xref="taxon:3352"
/clone="PC15H07"

/clone_lib="Pine Triplex pollen cone library"
/dev_stage="immature"
/lab_host="E. coli BM25.8"
/note="Organ: pollen cone; Vector: Lambda Triplex; Site: 1:
Sfil (A); Site: 2: Sfil (B); Immature pollen cones were
collected in the early spring, frozen and used for mRNA
isolation. The SMART-PCR method (Clontech) was used to
prepare a library from 1 ug total RNA, using the Lambda
Triplex vector. Plasmid subclones in pTriplex were
recovered by cre-lox excision in E. coli strain BM25.8 and
sequenced from the 5' end."

BASE COUNT 176 a 100 c 122 g 128 t
ORIGIN

alignment_scores:
Quality: 422.00 Length: 121
Ratio: 3.944 Gaps: 0
Percent Similarity: 88.430 Percent Identity: 68.595

alignment_block:
US-09-435-054-2 x AW981720 ..

Align seg 1/1 to: AW981720 from: 1 to: 526

```

9 AlaAlaGlyAlaGluAsnGlySerAlaAlaGlyGlyAlaAsnAsnGlyGI 25
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
162 GCTTCAATGCGACAAGATGCAAGCCCAACGCCAGCAGGATACAGACCTC 211
25 yAlaAlaGlnHisAlaAlaProAlaIleArgGluGlnAspArgLeuM 42
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
212 CGAGGATGGAGATAGAGAAAATTTGCTGTAGAGAGAACAGGACAGGTTT 261
42 etProIleAlaAsnValIleArgIleMetArgArgValLeuProAlaHis 58
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
262 TGCTATCGCAATGTCATCAGGATCATGAGAAAAGTCTCTCCACACACAT 311
59 AlalysIleSerAspAlaLysGluThrIleGlnGluCysValSerGI 75
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
312 GCGAAAAATTTCTCATGATGCAAGAGAAACAATACAGGAATGTGTTTCAG 361
75 uTyrIleSerPheIleThrGlyGluAlaAsnGluArgCysGlnArgGluG 92
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
362 ATTATATCAGTTTCATCAGTAGTGAAGCAATGACCGGTGCCAAAAGGAAC 411
92 lnArgLysThrIleThrAlaGluAspValLeuTrpAlaMetSerArgLeu 108
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
412 AAAGGAGAGACTATCATCTGCCGAGGATGTTTATGCGCCATGAATAAGTT 461
109 GlyPheAspTyrValGluProLeuGlyAlaTyrLeuHisArgTyrAr 125
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
462 GGTTTTGACAAATTCATGGATCCATTCAGTCTTACCTTCAGAGATATAG 511
125 gGluPheGluGly 129
|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
512 AGGTATTGAGGGT 524

```

seq_name: gb_est28:AL509098

seq_documentation_block: 441 bp mRNA EST 04-JAN-2001
LOCUS AL509098
DEFINITION Hordeum vulgare Barke developing caryopsis (3.-15.DAP)
Hordeum vulgare cDNA clone HV10L07V 5', mRNA sequence.

ACCESSION AL509098
VERSION AL509098.1 GI:12035601
KEYWORDS EST.
SOURCE barley.
ORGANISM Hordeum vulgare

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae
; Triticeae; Hordeum.
1 (bases 1 to 441)
REFERENCE
AUTHORS Michalek, W., Weschke, W., Pleissner, K.-P. and Graner, A.

[illegible]

Pharmaceuticals, Inc.
 4633 World Parkway Circle, St. Louis, MO 63134, USA
 Tel: 877-577-2733
 Fax: 314-427-3324
 Email: service@genomesystems.com.

FEATURES
 source Location/Qualifiers
 1..575
 /organism="Arabidopsis thaliana"
 /db_xref="taxon:3702"
 /clone="70150287"
 /clone_lib="A. thaliana, Ohio State clone set"
 /note="cDNA library was made from selected clones from the Arabidopsis thaliana Ohio State clone set."
 BASE COUNT 154 a 116 c 177 g 118 t 10 others
 ORIGIN

```

alignment_scores:
    Quality: 406.50      Length: 186
    Ratio: 2.863         Gaps: 4
    Percent Similarity: 76.344   Percent Identity: 46.237

alignment_block:
US-09-435-054-2 x AI995140 ..

Align seg 1/1 to: AI995140 from: 1 to: 575

18 ALAglyGlyAlaAsnAsnclyGlyAlaLaGlnGlnHisAlaAProAL 34
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
11 TCCGGTGGAGGCCAAACCGGAACACCAGAACGCAGTCCTCCTTTGTC 60
: |||||:|||||:|||||:|||||:|||||:|||||:|||||:
34 alleArgGLuGlnAspArgLeuMetProIleAlaAsnValIleArGlLem 51
: |||||:|||||:|||||:|||||:|||||:|||||:|||||:
61 TCCAAGAGAGCAAGACAGGTCTTCCCGATCGCTAACGTCAGCCGGATCA 110
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
51 etArgArgValLeuProAlaHisAlaLysIleSerAspAlaLysGlu 67
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
111 TGAAGAAGGCCCTTGCCCGCAACGCCCAAGATCTCTAAAGATGCCAAGAG 160
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
68 ThrILeGlnGluCyValSerLutTyrlleSerPheIleThrGLyGLuAL 84
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
161 ACGATGCAGGAGTGTGTTCCGAGTTCATCAGTTCGTCACGGGAGAGC 210
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
84 aAsnGluArgCysGlnArgGluGlnArgLysThrIleThraLaGluAspv 101
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
211 ATCTGATAGTGTACAGAGGAGAGAGAGGAGACGATCACGGAGACGATT 260
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
101 aLeuTrpAlaMetSerArgLeuGlyPheAspAspTyrValcluProLeu 117
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
261 TGCTCTGGCTATGACTACTCTAGCTTTTGGAGATTATGTTGAGCCATTG 310
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
118 GlyAlaTyrrLeuHisArgTyrrArgLupheGluGlyAspAlaArgGlyVa 134
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
311 AAAGTTTACTTGCAGAGGTTTAGGGAGATCGAAGGGGAGAGGACTTGGACT 360
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
134 lGlyLeuValProGlyAlaAlaProSerArgGlyGlyAspHisHisProH 151
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
361 AGGGAG.....GCACAGACTGGTGGTGAGGTCGGAGAGC 394
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
151 IsSerMetSerProAlaAlaMetLeuLysSerArgGlyProValSerGly 167
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
395 ATCAGAGAGATGCTCTCGGA.....GATGGCGGGTGGGTCTACGGT 435
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
168 AlaAlaMet...LeuProHisHisHisHisHisHisAspMetGlnMeHi 183
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
436 NNGTNGTGGGATCATGATATCACCACATCATCAGTTTCTTCCACGACA 485
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
183 sAlaAlaMetTyrrGlyThrAla.....V 192
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
486 GACCATATGATTATGAGCCACANNGTGGCGTACGACAGCTNANGTNAGC 535
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
192 alProPro 194
:::|||||

```

```

536 TGCTCCG 543
seq_name: gb_est85:BF263449
seq_documentation_block:
LOCUS BF263449 595 bp mRNA EST 09-MAR-2001
DEFINITION HV_CEA0006M10f Hordeum vulgare seedling green leaf EST library
HVCDA0004 (Erysiphe infected & control) Hordeum vulgare cDNA clone
HV_CEA0006M10f, mRNA sequence.
ACCESSION BF263449
VERSION BF263449.2 GI:13260832
KEYWORDS EST.
SOURCE barley.
ORGANISM Hordeum vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae
; Triticeae; Hordeum.
REFERENCE 1 (bases 1 to 595)
AUTHORS Wing,R., Close,T.J., Kleinhofs,A., Wise,R., Begum,D., Frisch,D., Yu
,Y., Anderson,H., Dale,J., Henry,D., Kernodle,S., Palmer,M., Rambo
,T., Saski,C., Schwartzbeck,J., Simmons,J., Choi,D.W., Main,D. and
Wood,T.
TITLE Development of a genetically and physically anchored EST resource
for barley genomics
JOURNAL Unpublished (2000)
COMMENT On Nov 17, 2000 this sequence version replaced gi:11194443.
Contact: Wing RA
Clemson University Genomics Institute
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Seq primer: AATTAACTCTCACTAAAGGG
High quality sequence stop: 594.
Location/Qualifiers
1..595
/organism="Hordeum vulgare"
/cultivar="C116155 (M1a13)"
/db_xref="taxon:4513"
/clone="HV_CEA0006M10f"
/clone_lib="Hordeum vulgare seedling green leaf EST
library HVCDA0004 (Erysiphe infected & control)"
/tissue_type="seedling green leaf"
/lab_host="TJG121"
/note="Vector: lambdaZAP; Site_1: EcoRI; Site_2: XhoI"
BASE COUNT 133 a 199 c 173 g 87 t 3 others
ORIGIN

alignment_scores:
Quality: 405.00 Length: 184
Ratio: 3.140 Gaps: 4
Percent Similarity: 70.109 Percent Identity: 46.196

alignment_block:
US-09-435-054-2 x BF263449 ..
Align seg 1/1 to: BF263449 from: 1 to: 595

14 ASnGlySerAlaAlaGlyGlyAlaAsnAnGlyGlyAlaAlaGlnClnHi 30
:.....:.....:.....:.....:.....:.....:
79 GACAAGACATCCAGTGGCGGAGCAACAGCGGCGG.....GAGCT 119
:.....:.....:.....:.....:.....:.....:
30 sAlaAlaPrAlaAlaArgGluGlnAspArgLeuMetProIleAlaAsnV 47
:.....:.....:.....:.....:.....:.....:
120 CTGCTCCCG.....CGGAGGAGGAGCGCTTCTGCGCCATCGCCCAAGC 163
:.....:.....:.....:.....:.....:.....:
47 allleArgIleMetArgValLeuProAlaAlaHisAlaLysIleSerAsp 63
|| .....:.....:.....:.....:.....:
164 TCAGCGCGATCATGAAGAGCGCTCCGCGCAACGCCAAGATCAGCAAA 213
:.....:.....:.....:.....:.....:
64 AspAlaLysGluThrIleGlnGluCysValSerGluTyrIleSerPheI 80
:.....:.....:.....:.....:.....:

```



```

126 uPheGluGlyAspAlaArgGlyValGlyLeuValProGlyAlaAlaProS 143
|:::|||||:::||||| ||| |||:::|
480 GATGGAGGTGACACTAAGGATCAGCA.....AAGGCGGGGAATCGA 523

143 erArgGlyGlyAspHisHisProHisSerMetSer 154
|| |||:::|||||:::|||||
524 GTGCAAGAGAGATGGTCAACCAAGCCCAAGTGTCT 558

```


PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 09/026,221
; FILING DATE: 19-FEB-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/804,534
; FILING DATE: 21-FEB-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Einhorn, Gregory P.
; REGISTRATION NUMBER: 38,440
; REFERENCE/DOCKET NUMBER: 023070-077611US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 627 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..627
; OTHER INFORMATION: /product= "LECl"
US-09-103-478-1

Query Match 15.1%; Score 177; DB 4; Length 627;
Best Local Similarity 68.1%; Pred. No. 2.2e-27;
Matches 246; Conservative 0; Mismatches 115; Indels 0; Gaps 0;
QY 97 ccggcgaggagatggtcgtcgccggcgccgcaacaatggcgcgctgctcagcagc 156
Db 5 CCAGCTAGTCATAGTAGCCGCGCCGCTGACAAAGAACATGGTATCGTCCAGCAGC 64
QY 157 atcgcgccgagcgatccgagcagcagcagcagcagcagcagcagcagcagcagc 216
Db 65 AACACACATGTGTGGCTCGTGAGCAAGAACCAATACATGCCAATCGCAAGCTCATAAGAA 124
QY 217 tcatcgccgctgctcgtccgagcagcagcagcagcagcagcagcagcagcagcagc 276
Db 125 TCATCGCTAAACCTTACCGTCTCAGCCCAAAATCTCTGACGAGCGCAAGAAACGATTC 184
QY 277 aggaagtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtc 336
Db 185 AAGAAATGTCTCCGAGTACATCAGCTTCGTGACCGGTGAAGCAACGAGCGTTGCCAAC 244
QY 337 gggagcagcgaagaccatccagcagcagcagcagcagcagcagcagcagcagcagc 396
Db 245 GTGAGCAACGTAGACCATTAACCTGCTGAAGATATCTTTGGGGCTATGAGCAAGCTTGGGT 304
QY 397 tcgagcactcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtc 456
Db 305 TCGATAACTAGTGGAGCCCTCCACCGTGTTCATTAACCGGTACCGTACCGTACGATAGAGACCG 364
QY 457 a 457
Db 365 A 365

RESULT 4
US-09-319-989-3
; Sequence 3, Application US/09319989
; Patent No. 6190914
; GENERAL INFORMATION:
; APPLICANT: Grivell, Leslie A.
; APPLICANT: Teixeira De Mattos, Maarten J.
; APPLICANT: Blom, Jolanda
; TITLE OF INVENTION: METHODS FOR MODULATING METABOLIC PATHWAYS OF
; TITLE OF INVENTION: MICRO-ORGANISMS AND MICRO-ORGANISMS OBTAINABLE BY SAID
; FILE REFERENCE: 24615-20123.00
; CURRENT APPLICATION NUMBER: US/09/319,989

; CURRENT FILING DATE: 1999-06-14
; EARLIER APPLICATION NUMBER: PCT/NL97/00688
; EARLIER FILING DATE: 1997-12-12
; EARLIER APPLICATION NUMBER: EPO 96203520
; EARLIER FILING DATE: 1996-12-12
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 835
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: DNA construct
; OTHER INFORMATION: encoding HAP3
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (201)..(632)
US-09-319-989-3

Query Match 6.7%; Score 78.4; DB 4; Length 835;
Best Local Similarity 53.2%; Pred. No. 1.2e-07;
Matches 166; Conservative 0; Mismatches 146; Indels 0; Gaps 0;
QY 119 ggcggcgccgcccacaacatggcgcgctcgtcagcagcagcagcagcagcagcagcagc 178
Db 251 ggagagcgggtggaacgctagctccagcgcgaggttgcagcaaatctccacgctaagaga 310
QY 179 gcaggaccggctgagtcgagcagcagcagcagcagcagcagcagcagcagcagcagc 238
Db 311 gcaggacagatggtaccatcaacaatgtagcgagcagcagcagcagcagcagcagcagc 370
QY 239 gcacgcagagatctcgagcagcagcagcagcagcagcagcagcagcagcagcagcagc 298
Db 371 gagtgtgaagtgatgaaagatgcgaagagtgcatgcaggagtggtcagtgagctcat 430
QY 299 cagcttcacacgggggagggccacagcagcagcagcagcagcagcagcagcagcagcagc 358
Db 431 tcttttggatagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 490
QY 359 cgccgagcagcgtgctgtgggcatgagcgcctcgctcgctcgctcgctcgctcgctcgctcg 418
Db 491 cggggaagacattctcatatcatctgacgccttagattcgagaactatgcagaggtgtt 550
QY 419 cggcgccctacct 430
Db 551 gaaaatctactt 562

RESULT 5
US-07-977-434-5
; Sequence 5, Application US/07977434
; Patent No. 5466591
; GENERAL INFORMATION:
; APPLICANT: Gelfand, David H.
; APPLICANT: Abramson, Richard D.
; TITLE OF INVENTION: 5' TO 3' EXONUCLEASE MUTATIONS OF
; TITLE OF INVENTION: THERMOSTABLE DNA POLYMERASES
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hoffmann-La Roche Inc.
; STREET: 340 Kingsland Street
; CITY: Nutley
; STATE: New Jersey
; ZIP: 07110-1199
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Macintosh
; OPERATING SYSTEM: 7
; SOFTWARE: WordPerfect 2.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/977,434
; FILING DATE:

[illegible]


```
; ANTI-SENSE: NO
US-08-457-342-6

Query Match      4.8%; Score 56; DB 1; Length 28958;
Best Local Similarity 48.2%; Pred. No. 0.0056;
Matches 158; Conservative 0; Mismatches 170; Indels 0; Gaps 0;

QY 356 caccgagagagctgtgtggccatgagcgccctcgcttcgagactacgtcgagcc 415
Db 16356 CGCCCTCTTTGCGGTCTATGCTCTCCCTGGCCGCCCTCTGCGCTCGCTCGGCGTCGAGCC 16415
QY 416 gctcgccgctactccacgcctaccgcaggttcgaggcgagcgcgcggtcggtc 475
Db 16416 CGCGCGCGTCTGCGGCACAGCCAGGCGGAGATAGCGCGCGCTTCGTCGACGGCGCTCT 16475
QY 476 cgtccggggggcgcccccacgcgcggcgagcgagccaccaccaccacccacgc 535
Db 16476 CTCCTCGAGGAGCGCGCGCATCGCGCCCTGCGCAGCAAGCGTCACCGCTCGCC 16535
QY 536 ggcgatctcaagtcccgcgggcgagtcctccgagcgccatctaccgcaccacacca 595
Db 16536 GCGAACGGGATGCGCGCCCTCGAGCTCGCGCTCGGAGCTCGGAGCTCGGAGCTCG 16595
QY 596 ccaccacagacatgcagatgcacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 655
Db 16596 CTGGGGCGACAGGCTCTCCATCGCGCGCTCAACAGCCCGGCGGCGCTGATC 16683
QY 656 cgggctctcaccacggcggttcctc 683
Db 16656 CGAGCGCGCGCGCTCGACGCGCTGATC 16683

RESULT 11
US-08-457-646A-6
; Sequence 6, Application US/08457646A
; Patent No. 5679560
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James M.
; APPLICANT: Beck, James Joseph
; APPLICANT: Hill, Dwight Steven
; APPLICANT: Ryals, John Andrew
; APPLICANT: Gaffney, Thomas Deane
; APPLICANT: Lam, Stephen Ting
; APPLICANT: Hammer, Phillip E.
; APPLICANT: Uknes, Scott Joseph
; TITLE OF INVENTION: Genes for the synthesis of
; TITLE OF INVENTION: antipathogenic substances
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ciba-Geigy Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/457,646A
; FILING DATE: 01-JUN-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/457,205
; FILING DATE: 01-JUN-1995
; APPLICATION NUMBER: 08/258,261
; FILING DATE: 08-Jun-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Elmer, James Scott
```

```
; REGISTRATION NUMBER: 36,129
; REFERENCE/DOCKET NUMBER: CGC 1506/CIP3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8614
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 28958 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-457-646A-6

Query Match      4.8%; Score 56; DB 1; Length 28958;
Best Local Similarity 48.2%; Pred. No. 0.0056;
Matches 158; Conservative 0; Mismatches 170; Indels 0; Gaps 0;

QY 356 caccgagagagctgtgtggccatgagcgccctcgcttcgagactacgtcgagcc 415
Db 16356 CGCCCTCTTTGCGGTCTATGCTCTCCCTGGCCGCCCTCTGCGCTCGCTCGGCGTCGAGCC 16415
QY 416 gctcgccgctactccacgcctaccgcaggttcgaggcgagcgcgcggtcggtc 475
Db 16416 CGCGCGCGTCTGCGGCACAGCCAGGCGGAGATAGCGCGCGCTTCGTCGACGGCGCTCT 16475
QY 476 cgtccggggggcgcccccacgcgcggcgagcgagccaccaccaccacccacgc 535
Db 16476 CTCCTCGAGGAGCGCGCGCATCGCGCCCTGCGCAGCAAGCGTCACCGCTCGCC 16535
QY 536 ggcgatctcaagtcccgcgggcgagtcctccgagcgccatctaccgcaccacacca 595
Db 16536 GCGAACGGGATGCGCGCCCTCGAGCTCGCGCTCGGAGCTCGGAGCTCGGAGCTCG 16595
QY 596 ccaccacagacatgcagatgcacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 655
Db 16596 CTGGGGCGACAGGCTCTCCATCGCGCGCTCAACAGCCCGGCGGCGCTGATC 16683
QY 656 cgggctctcaccacggcggttcctc 683
Db 16656 CGAGCGCGCGCGCTCGACGCGCTGATC 16683

RESULT 12
US-08-458-076A-6
; Sequence 6, Application US/08458076A
; Patent No. 5698425
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James M.
; APPLICANT: Beck, James Joseph
; APPLICANT: Hill, Dwight Steven
; APPLICANT: Ryals, John Andrew
; APPLICANT: Gaffney, Thomas Deane
; APPLICANT: Lam, Stephen Ting
; APPLICANT: Hammer, Phillip E.
; APPLICANT: Uknes, Scott Joseph
; TITLE OF INVENTION: Genes for the synthesis of
; TITLE OF INVENTION: antipathogenic substances
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ciba-Geigy Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
```

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/458,076A
FILING DATE: 01-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/457,205
FILING DATE: 01-JUN-1995
APPLICATION NUMBER: 08/258,261
FILING DATE: 08-Jun-1994
ATTORNEY/AGENT INFORMATION:
NAME: Elmer, James Scott
REGISTRATION NUMBER: 36,129
REFERENCE/DOCKET NUMBER: CGC 1506/CIP3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8614
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 28958 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-458-076A-6

Query Match 4.8%; Score 56; DB 1; Length 28958;
Best Local Similarity 48.2%; Pred. No. 0.0056;
Matches 158; Conservative 0; Mismatches 170; Indels 0; Gaps 0;

QY 356 caccgcgcagagctgtgtggtgcatgagccgctcgttgcagactacgtcgagcc 415
Db 16356 CGCCCTCTTTCCGCTCATGCTCTCCCTGCGCCCTCTGCGCTCGCTCGCGCTCGAGCC 16415

QY 416 gctcggcgctacccacccgctaccgagcttcgagggcgagcgcgcgctcgagct 475
Db 16416 CGCCCGCTGCTCGGCGCACACAGCGAGGGCGGAGATAGCCGCGCTCTGCTCGAGCGCTCT 16475

QY 476 cgtcccgggggcccccattcgccgagcaccaccccgccatccatgtcgcagc 535
Db 16476 CTCCCTCGAGACGCGCGCGATCGCGCCCTCGCGAGCAAGCGTCACACCGTCCG 16535

QY 536 ggcagatcgaatgagatgacgcgcgcagcttcgagggcgagcgcgcgctcgagct 595
Db 16536 GGCAACGGGATGCGCGCGCTCGAGCTCGGCGCTCGAGCTCCAGACCTACCTCGCTCC 16595

QY 596 ccaccagacatgagatgacgcgcgcagcttcgagggcgagcgcgcgctcgagct 655
Db 16596 CTGGGCGACAGGCTCTCCATCGCGCGCTCAACAGCGCCGAGGGCGAGCGCTCGATCCGG 16655

QY 656 cggcgctcctaccacggcggttcctc 683
Db 16656 CGAGCGCGCGCGCTCGAGCGCTGATC 16683

RESULT 13
US-08-764-233A-4
Sequence 4, Application US/08/64233A
Patent No. 5716849
GENERAL INFORMATION:
APPLICANT: Ligon, James M.
APPLICANT: Schupp, Thomas
APPLICANT: Beck, James J.
APPLICANT: Hill, Dwight S.
APPLICANT: Neff, Snezana
APPLICANT: Ryals, John A.
TITLE OF INVENTION: Genes For The Biosynthesis Of Soraphen
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:

ADDRESSEE: Ciba-Geigy Corporation
STREET: 520 White Plains Road, P.O. Box 2005
CITY: Tarrytown
STATE: NY
COUNTRY: USA
ZIP: 10591
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/764,233A
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/729,214
FILING DATE: 09-OCT-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/258,261
FILING DATE: 08-JUN-1994
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: 1506/CIP6
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919) 541-8587
TELEFAX: (919) 541-8689
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 28958 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Sorangium cellulosum
IMMEDIATE SOURCE:
CLONE: p98/1
US-08-764-233A-4

Query Match 4.8%; Score 56; DB 1; Length 28958;
Best Local Similarity 48.2%; Pred. No. 0.0056;
Matches 158; Conservative 0; Mismatches 170; Indels 0; Gaps 0;

QY 356 caccgcgcagagctgtgtggtgcatgagccgctcgttgcagactacgtcgagcc 415
Db 16356 CGCCCTCTTTCCGCTCATGCTCTCCCTGCGCCCTCTGCGCTCGCTCGAGCC 16415

QY 416 gctcggcgctacccacccgctaccgagcttcgagggcgagcgcgcgctcgagct 475
Db 16416 CGCCCGCTGCTCGGCGCACACAGCGAGGGCGGAGATAGCCGCGCTCTGCTCGAGCGCTCT 16475

QY 476 cgtcccgggggcccccattcgccgagcaccaccccgccatccatgtcgcagc 535
Db 16476 CTCCCTCGAGACGCGCGCGATCGCGCCCTCGCGAGCAAGCGTCACACCGTCCG 16535

QY 536 ggcagatcgaatgagatgacgcgcgcagcttcgagggcgagcgcgcgctcgagct 595
Db 16536 GGCAACGGGATGCGCGCGCTCGAGCTCGGCGCTCGAGCTCCAGACCTACCTCGCTCC 16595

QY 596 ccaccagacatgagatgacgcgcgcagcttcgagggcgagcgcgcgctcgagct 655
Db 16596 CTGGGCGACAGGCTCTCCATCGCGCGCTCAACAGCGCCGAGGGCGAGCGCTCGATCCGG 16655

QY 656 cggcgctcctaccacggcggttcctc 683
Db 16656 CGAGCGCGCGCGCTCGAGCGCTGATC 16683

RESULT 14
US-08-457-335A-6

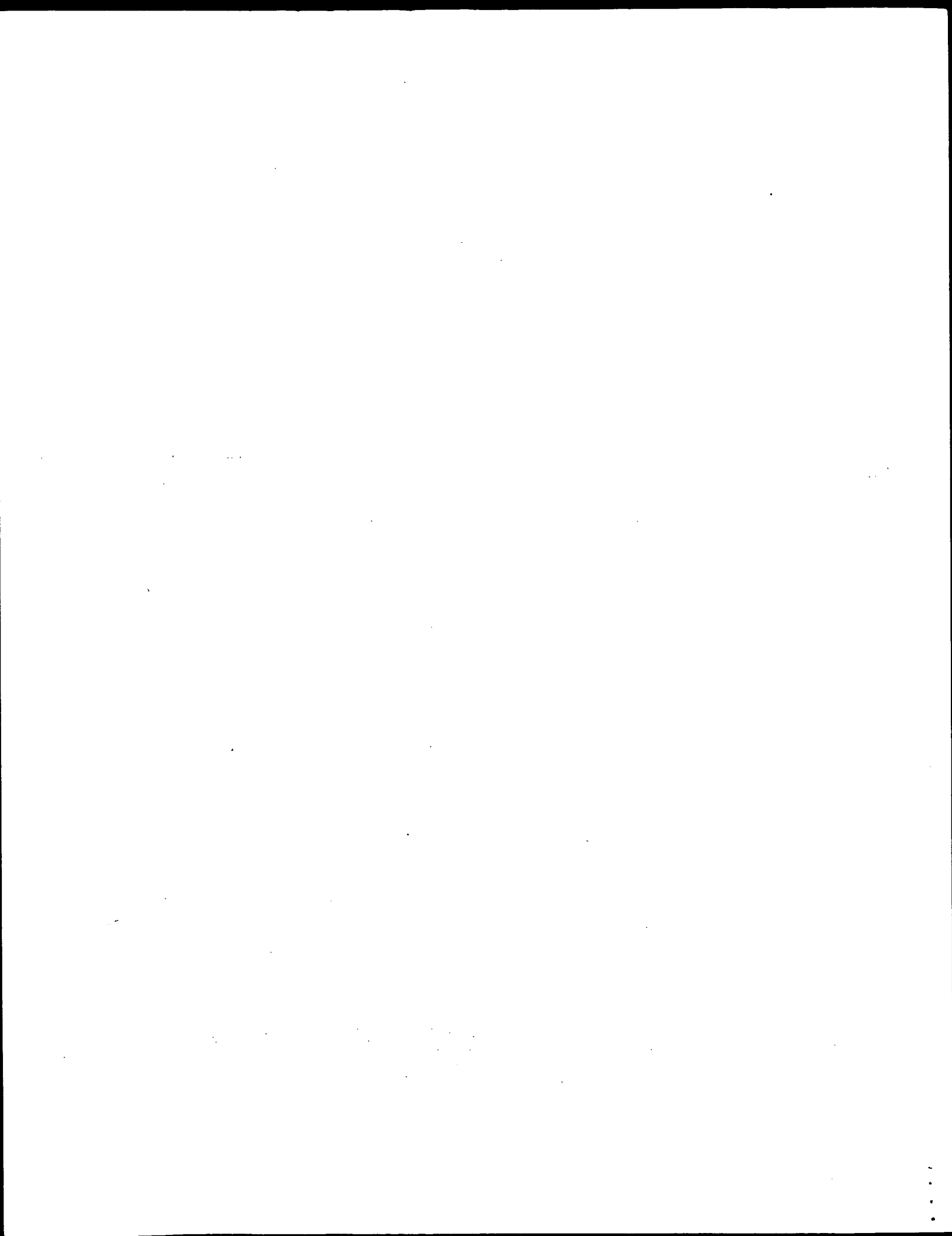
QY	356	caccgcccaggacgtgctgtggtggccatgagccgcctggcttcgagcactagctgagccc	415
Db	16356	CGCCCTCTTTGCGGTATGGTCTCCCTGGCCGCTCTGCGCTCGCTCGGGGTGAGCC	16415
QY	416	gtctggcgccctaacctccaccgctaccgcgaattctgaggggcagcgcgcgctcgggct	475
Db	16416	CGCGCGCGTCGTGCGCCACAGCCAGGCGAGATAGCGCGCGCTTCGTGCGAGGCGCTCT	16475
QY	476	cgtcccgggggcgccgccatctgcgcgcgcgagccaccgccacatccatctgcgcagc	535

```

Db 16476 CTCCCTCGAGGACGGCGCGCATCGCGCCCTGGCAGCAAGCGTCACACCGTCGCC 16535
QY 536 ggcgatgtcaagtccgcgggccaagtctccggagccgccaatgtaccgacaccacca 595
Db 16536 GGCAACGGGCATGGCGCGCGTCGAGCTCGGGCGCCTCCGACCTCCAGACCTACCTCGCTCC 16595
QY 596 ccaccacgacatgcagatgcacgcgcgcctgtacgggggaaacggccgtgccccccgcggc 655
Db 16596 CTGGGGCGACAGGCTCTCCATCGCGCGCGCTCAACAGCCCCAGGGCCACGCTCGTATCCGG 16655
QY 656 cgggctctctaccacgcgggttctc 683
Db 16656 CGAGCCCGCGCGCTCGACGCGCTGATC 16683

```

Search completed: October 25, 2001, 12:01:59
Job time: 4344 sec



OM of: US-09-435-054-2 to: Issued Patents_NA.* out_format : pfs

Date: Oct 25, 2001 2:25 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=frame+P2n.model -DEV=xlh
-Q/cgn2_1/USPTO_Spool/US09435054/runat_25102001-085523-6989/app_query.fasta_1.338
-DB=Issued Patents_NA -QFMT=fastap -SUFFIX=p2n.rni
-GAPOP=12.000 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000
-LOOPEXT=0.000 -CGAPOP=4.500 -CGAPEXT=0.050 -XGAPOP=10.000
-XGAPEXT=0.500 -FCGAPOP=6.000 -FCGAPEXT=7.000 -YGAPOP=10.000
-YGAPEXT=0.500 -DELOP=6.000 -DELEXT=7.000 -START=1
-MATRIX=blosom62 -TRANS=human40.cdi -LIST=45 -DICALIGN=200
-THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pfs -NORM=ext -MINLEN=0 -MAXLEN=200000000
-USER=US09435054.CGN1_1.77 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XUPXY -WAIT -THREADS=1

Search information block:

Query: US-09-435-054-2

Query length: 278

Database: Issued Patents_NA.*

Database sequences: 324599

Database length: 94655562

Search time (sec): 38.160000

score_list:

Sequence	Strd Orig	ZScore	Escore	Len	Documentation
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-103-478-1 +	494.00	676.81	4.4e-30	627	1
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-103-478-3 +	494.00	660.44	3.6e-29	3395	1
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-103-478-4 +	494.00	652.68	9.7e-29	7560	1
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-319-989-3 +	295.50	400.48	1.1e-14	835	1
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-758-662-4 +	167.50	228.50	4.1e-05	530	1
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-845-998-7 +	167.50	207.72	0.0006	4524	1
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-206-537-7 +	167.50	207.72	0.0006	4524	1
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-553-110-2 +	147.00	188.12	0.0073	1853	1
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-041-886-22 +	147.00	180.00	0.0207	4279	1
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-041-886-1 +	141.50	178.98	0.0236	2176	1
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-041-886-10 +	136.50	166.90	0.1110	3715	1
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-073-384-2 +	134.50	168.00	0.0964	2496	1
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-254-359A-2 +	134.50	168.00	0.0964	2496	1
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-483-043-2 +	134.50	168.00	0.0964	2496	1
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-481-238-2 +	134.50	168.00	0.0964	2496	1
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-471-066B-2 +	134.50	168.00	0.0964	2496	1
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-484-956-2 +	134.50	168.00	0.0964	2496	1
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-757-653-2 +	134.50	168.00	0.0964	2496	1
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-599-491-2 +	134.50	168.00	0.0964	2496	1
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-756-386-2 +	134.50	168.00	0.0964	2496	1
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-823-516-2 +	134.50	168.00	0.0964	2496	1
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-682-853A-2 +	134.50	168.00	0.0964	2496	1
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-759-038-2 +	134.50	168.00	0.0964	2496	1
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-758-314-2 +	134.50	168.00	0.0964	2496	1
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-781-803-7 +	134.00	165.07	0.1405	3147	1
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-694-078-7 +	134.00	165.07	0.1405	3147	1
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-058-260-7 +	134.00	165.07	0.1405	3147	1
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-864-038A-1 +	132.50	166.41	0.1183	2214	1
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-864-038A-2 +	132.50	166.41	0.1183	2214	1
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-864-038A-4 +	132.50	162.45	0.1966	3331	1
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-581-148C-30 +	129.00	162.45	0.1966	3331	1
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-911-853-30 +	128.50	159.21	0.2977	6343	1
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-479-409-30 +	128.50	159.21	0.2977	2634	1
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-911-853-29 +	128.50	159.21	0.2977	2634	1
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-479-409-29 +	128.50	140.80	3.16	17612	1
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-479-409-29 +	128.50	140.80	3.16	17612	1
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-083-351-3 +	125.50	159.56	0.2848	1659	1
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-083-352-3 +	125.50	159.56	0.2848	1659	1
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-083-351-1 +	125.50	151.16	0.8362	3946	1
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-083-352-1 +	125.50	151.16	0.8362	3946	1
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-804-227C-1 -	125.50	127.95	16.42	43280	1

/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-928-213B-7 - 124.50 153.53 0.6172 2681
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-690-473-1 + 124.50 149.05 1.10 4257
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-259-821A-1 + 124.50 149.05 1.10 4257
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-843-659-1 + 124.50 149.05 1.10 4257
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-458-568A-11 - 124.50 139.00 3.98 1200

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-103-478-1

seq_documentation_block:
; Sequence 1, Application US/09103478
; Patent No. 6235975
; GENERAL INFORMATION:
; APPLICANT: Harada, John
; APPLICANT: Lotan, Tamar
; APPLICANT: Ohno, Masa-aki
; APPLICANT: Goldberg, Robert B.
; APPLICANT: Fischer, Robert L.
; TITLE OF INVENTION: LEAFY COTYLEDON1 Genes and Their Uses
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/103,478
; FILING DATE: 24-JUN-1998
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 09/026,221
; FILING DATE: 19-FEB-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/804,554
; FILING DATE: 21-FEB-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Einhorn, Gregory P.
; REGISTRATION NUMBER: 38,440
; REFERENCE/DOCKET NUMBER: 023070-077611US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 627 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..627
; OTHER INFORMATION: /product= "LECI"

US-09-103-478-1

alignment_scores:

Quality: 494.00 Length: 249

Ratio: 3.146 Gaps: 9

Percent Similarity: 63.052 Percent Identity: 46.988

alignment_block:

US-09-435-054-2 x US-09-103-478-1 ..

Align seg 1/1 to: US-09-103-478-1 from: 1 to: 627

16 SerAlaAlaGlyAlaAsnGlyGlyAlaAlaGlnHisAlaA1 32

ADDRESS: Townsend and Townsend and Crew LLP
 STREET: Two Embarcadero Center, Eighth Floor
 CITY: San Francisco
 STATE: California
 COUNTRY: USA
 ZIP: 94111-3834
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/103,478
 FILING DATE: 24-JUN-1998
 CLASSIFICATION: 800
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 09/026,221
 FILING DATE: 19-FEB-1998
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/804,534
 FILING DATE: 21-FEB-1997
 ATTORNEY/AGENT INFORMATION:
 NAME: Einhorn, Gregory P.
 REGISTRATION NUMBER: 38,440
 REFERENCE/DOCKET NUMBER: 023070-077611US
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415) 576-0200
 TELEFAX: (415) 576-0300
 INFORMATION FOR SEQ ID NO: 3:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 3395 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 US-09-103-478-3

```

alignment_scores:      Length: 249
      Quality: 494.00      Caps: 9
      Ratio: 3.146
      Percent Similarity: 63.052
alignment_block:
US-09-435-054-2 x US-09-103-478-3 ..
Align seg 1/1 to: US-09-103-478-3 from: 1 to: 3395

16 SerAlaAlaGlyGlyAlaAsnAsnGlyGlyAlaAlaGlnGlnHisAlaAl 32
   ::::::::::::::::::::::::::::::::::::::::::::::
2020 GCCGCGCGCGGTGCAAGAACAATGTAATGCTGGTCCAGCAGCAACCACC 2069

32 aProAlaIleArgGluGlnAspArgLeuMetProIleAlaAsnValIleA 49
   ::::::::::::::::::::::::::::::::::::::::::::::
2070 ARGGTGGTCTGTGAGCAGACCAATACATACGATCGCAATCGCAACGTCNTAA 2119

49 rGileMetArgArgValLeuProAlaHisAlaLysIleSerAspAla 65
   ::::::::::::::::::::::::::::::::::::::::::::::
2120 GAATCATGCGTAAACCTTTACCGTCTCACGCCCAAAATCTCTGACGACGCC 2169

66 LysGluThrIleGlnGluCysValSerGluTyrIleSerPheIleThrGl 82
   ::::::::::::::::::::::::::::::::::::::::::::::
2170 AAAGAAACGATTTCAGAAATGTGTCTCCGAGTACATCAGCTTGTGACCGG 2219

82 yGluAlaAsnGluArgCysGlnArgGluGlnArgLysThrIleThrAlaG 99
   ::::::::::::::::::::::::::::::::::::::::::::::
2220 TGAAGCCAACGAGCGTGTGCCAAGTGGAGCAACGTAAGACCATAACTGCTG 2269

99 luAspValLeuTrpAlaMetSerArgLeuGlyPheAspAspTyrValGlu 115
   ::::::::::::::::::::::::::::::::::::::::::::::
2270 AAGATATCCTTTGGGGTATGAGCAAGCTTGGGTTCGATAACTACGTGGAC 2319

116 ProLeuGlyAlaTyrLeuHisArgTyrArgGluPheGluGlyAspAlaR 132

```

66 LysGluThrIleGlnGluCysValSerGluTyrIleSerPheIleThrG1 82
 |||||
 2170 AAAGAAACGATTCAAGAAATGTCTCCGAGTACATCAGCTTGGTGACCGG 2219
 |||||
 82 yGluAlaAsnGluArgCysGlnArgGluGlnArgLysThrIleThrAlaG 99
 |||||
 2220 TGAAGCCAACGAGCGTGTGCCAGCTGAGCAACGTAAGACCATAACTGCCTG 2269
 |||||
 99 luAspValLeuTrpAlaMetSerArgLeuGluIypheAspAspTyrValGlu 115
 |||||
 2270 AGATATCTCTGGGCTATGAGCAACGCTTGGTTCGATAACTACGTGGAC 2319
 |||||
 116 ProLeuGlyAlaTyrLeuHisArgTyrArgGluuPheGluGlyAspAlaAr 132


```

||||| ..... ||||||| ..... ||| ||| |||
2320 CCCCTACCGTGTTCATTACCGGTACCGTGAGATAGACCGAT...CG 2366
132 gGlyValGlyLeuValProGlyAlaAlaProSerArgGlyGlyAspHisH 149
||||| ..... ||||||| ..... AGAGGT..... 2385
2367 TGGTTCGCACTT..... 2385
149 isProHisSerMetSerProAlaAlaMetLeuLysSerArgGlyProVal 165
||||| ..... ||||||| ..... 2386
2386 ..... GAGCCACCGTCGTTGAGACAAACC..... 2409
166 SerGlyAlaAlaMetLeuProHisHisHisHisHisAspMetGlnMe 182
2409 ..... 2409
182 thisAlaAlaMetTyGlyGlyThrAlaValProProAlaGlyProp 199
||||| ..... 2410
2410 ..... TATGGAGGAATGTTAT..... 2427
199 roHisHisGlyGlyPheLeuMetProHisProGlnGlySerSerHisTyr 215
||||| ..... CACGGCCCATCTCATGGC 2451
2428 ..... GGGTTT..... 2451
216 Leu.....ProTyAlaTyrGluProThrTyrGlyGlyGluH 228
||||| ..... GGTATGTT 2486
2452 CTACCTCCTCCGGGTCCTTATGTTAT..... 2486
228 sAlaMetAlaAlaTyrTyGlyGlyAlaAlaTyrAlaProGlyAsnGlyG 245
||||| ..... 2487
2487 GGACCAATCATGTTATGGAGGTGTCGTACTACCAA...AAGCGGT 2533
245 lySerGlyAspGlySerGlySerGlyGlyGlyGlyGlyGlySerAlaSer 260
||||| ..... 2534
2534 CGTCGGGTCAAGATGATCAAGTGTGGTGGTGGTTCGTCCTTC 2580

```

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:us-09-103-478-4

seq_documentation_block:

; Sequence 4, Application US/09103478

; Patent No. 6235975

; GENERAL INFORMATION:

; APPLICANT: Harada, John

; APPLICANT: Lotan, Tamar

; APPLICANT: Onto, Masa-aki

; APPLICANT: Goldberg, Robert B.

; APPLICANT: Fischer, Robert L.

; TITLE OF INVENTION: LEAFY COTYLEDON1 Genes and Their Uses

; NUMBER OF SEQUENCES: 29

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Townsend and Townsend and Crew LLP

; STREET: Two Embarcadero Center, Eighth Floor

; CITY: San Francisco

; STATE: California

; COUNTRY: USA

; ZIP: 94111-3834

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/103,478

; FILING DATE: 24-JUN-1998

; CLASSIFICATION: 800

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 09/026,221

; FILING DATE: 19-FEB-1998

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/804,534

; FILING DATE: 21-FEB-1997

; ATTORNEY/AGENT INFORMATION:

; NAME: Einhorn, Gregory P.

; REGISTRATION NUMBER: 38,440

; REFERENCE/DOCKET NUMBER: 023070-077611US

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (415) 576-0200

; TELEFAX: (415) 576-0300

; INFORMATION FOR SEQ ID NO: 4:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 7560 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

; US-09-103-478-4

alignment_scores:

Quality: 494.00 Length: 249

Ratio: 3.146 Gaps: 9

Percent Similarity: 63.052 Percent Identity: 46.988

alignment_block:

US-09-435-054-2 x US-09-103-478-4 ..

Align seg 1/1 to: US-09-103-478-4 from: 1 to: 7560

16 SerAlaAlaGlyGlyAlaAlaAsnGlyGlyAlaAlaGlnGlnHisAlaAl 32

4449 GCCGGCGCGGTGACAAAGCAACATGTTATCTGGTCCAGCAGCAACACC 4498

32 aProAlaIleArgGluGlnAspArgLeuMetProIleAlaAsnValIleA 49

4499 ATGTGTGTGCTCGTGAGCAAGCAACATACATGCCAATCGCAACGTCATAA 4548

49 rGleMetArgValLeuProAlaHisAlaLysIleSerAspAla 65

4549 GAATCATCGTAAACCTTACCCTCTCACGCCAAATCTCTGACGACGCC 4598

66 LysGluThrIleGlnGluCysValSerGluTyrIleSerPheIleThrGl 82

4599 AAGAAACGATTCAAGATGTCTCCGAGTACATCAGCTTCATGACCGG 4648

82 yGluAlaAsnGluArgCysGlnArgGluGlnArgLysThrIleThrAlaG 99

4649 TGAAGCCCAACGAGCGTGGCCCAACGTGAGCAACGTAAGACCACTACTGCTG 4698

99 luAspValLeuTyrAlaMetSerArgLeuGlyPheAspArgTyrValGlu 115

4699 AAGATATCTTTGGGCTATGAGCAAGCTTGGTTCGATTAAGTACGTCGAC 4748

116 ProLeuGlyAlaTyrLeuHisArgTyrArgGluPheGluGlyAspAlaAr 132

4749 CCCCTCACCGTGTTCATTAAACCGGTACCGTGAGATAGACCGGAT...CG 4795

132 gGlyValGlyLeuValProGlyAlaAlaProSerArgGlyGlyAspHisH 149

4796 TGGTTCGCACTT.....AGAGGT..... 4814

149 isProHisSerMetSerProAlaAlaMetLeuLysSerArgGlyProVal 165

4815GAGCCACCGTCGTTGAGACAAACC..... 4838

166 SerGlyAlaAlaMetLeuProHisHisHisHisHisHisAspMetGlnMe 182

4838 4838

182 thisAlaAlaMetTyGlyGlyThrAlaValProProAlaGlyProp 199

4839TATGGAGGAATGTTAT..... 4856

199 roHisHisGlyGlyPheLeuMetProHisProGlnGlySerSerHisTyr 215

4857GGGTTT.....CACGGCCCATCTCATGGC 4880

216 Leu.....ProTyAlaTyrGluProThrTyrGlyGlyGluH 228

583

464

422 TGAGCTCATTTCTTTTGTGACATAGCGAGCCAGCATCGATCGCGTCTGCTG 471
91 luGlnArgLysThrIleThrAlaGluAspValLeuTrpAlaMetSerArg 107
472 ACAAGAAGACGATACCGGGAGACATCTTCATATCATTCATTCGACGCC 521
108 LeuGlyPheAspAspTyrValGluProLeuGlyAlaTyrLeuHisArgTy 124
522 TTAGATTGAGAACTATGACAGAGGTGTGAAATCTACTTCTGGCTAATA 571
124 rArgGlu 126
572 CAGCAA 578

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-758-662-4
seq_documentation_block:
; Sequence 4, Application US/08758662
; Patent No. 6114150
; GENERAL INFORMATION:
; APPLICANT: Weissman, Sherman
; APPLICANT: Baskaran, Namadev
; TITLE OF INVENTION: AMPLIFICATION OF NUCLEIC ACIDS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 701 Fifth Avenue, 6300 Columbia Center
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/758,662
; FILING DATE: 29-NOV-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 6114150tenburg Ph.D., Carol
; REGISTRATION NUMBER: 39,317
; REFERENCE/DOCKET NUMBER: 390036.402C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 530 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-758-662-4

alignment_scores:
Quality: 167.50 Length: 169
Ratio: 1.971 Gaps: 11
Percent Similarity: 50.296 Percent Identity: 35.503
alignment_block:
US-09-435-054-2 x US-08-758-662-4 ..
Align seg 1/1 to: US-08-758-662-4 from: 1 to: 530
125 ArgGluPheGluGlyAspAlaArgGlyValGly.....Le 136
67 CGAGCGCTGCTGGCGCGCGGAGCGCTGGCGCGCGTGACATCGCGG 116
136 uValProGlyAlaAlaProSer...ArgGlyGlyAspHisHisProHis 152
117 TGTCAGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 166

4881 CTACCTCCCTCGGGTCTTATGTTAT.....GGTATGTT 4915
228 sAlaMetAlaAlaTyrGlyGlyAlaAlaTyrAlaProGlyAsnGlyG 245
4916 GGACCAATCCATGTTATGGGAGGTGGTGGTACTACCAA...AACGGGT 4962
245 lySerGlyAspGlySerGlySerGlyGlyGlyGlyGlySerAlaSer 260
4963 CGTCGGGTCAAGATGAATCAAGTGTGGTGGTGGTCTTCGCTTCC 5009
seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-319-989-3

seq_documentation_block:
; Sequence 3, Application US/09319899
; Patent No. 6190914
; GENERAL INFORMATION:
; APPLICANT: Grivell, Leslie A.
; APPLICANT: Teixeira De Mattos, Maarten J.
; APPLICANT: Blom, Jolanda
; TITLE OF INVENTION: METHODS FOR MODULATING METABOLIC PATHWAYS OF
; TITLE OF INVENTION: MICRO-ORGANISMS AND MICRO-ORGANISMS OBTAINABLE BY SAID
; FILE REFERENCE: 24615-20123.00
; CURRENT APPLICATION NUMBER: US/09/319,989
; CURRENT FILING DATE: 1999-06-14
; EARLIER APPLICATION NUMBER: PCT/NL97/00688
; EARLIER FILING DATE: 1997-12-12
; EARLIER APPLICATION NUMBER: EPO 96203520
; EARLIER FILING DATE: 1996-12-12
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 835
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: DNA construct
; OTHER INFORMATION: encoding HAP3
; NAME/KEY: CDS
; LOCATION: (201)..(632)
US-09-319-989-3

alignment_scores:
Quality: 295.50 Length: 119
Ratio: 3.283 Gaps: 2
Percent Similarity: 75.630 Percent Identity: 50.420
alignment_block:
US-09-435-054-2 x US-09-319-989-3 ..
Align seg 1/1 to: US-09-319-989-3 from: 1 to: 835
8 ProAlaAlaGlyAlaGluAsnGlySerAlaAlaGlyGlyAlaAsnG 24
237 CCAGAGGATCTACGAGAGACGGT.....GGAACGCTAGCTCCAG 277
24 yGlyAlaAlaGlnGlnHisAlaAlaProAlaIleArgGluGluAspArgL 41
278 CGGCAGTTTGCAGCAA.....ATTTCACGCTTAAGAGACGAGACAGAT 321
41 euMetProIleAlaAsnValIleMetArgArgValLeuProAla 57
322 GGCTACCATCAACATGTACGCGACTCATGAAGATACTCTCCACCG 371
58 HisAlaLysIleSerAspAlaLysGluThrIleGlnGluCysVal 74
372 AGTGCTAAGGTATCGAAAGATCGCAAGAGAGTGCATGACGAGGTGTGTCAG 421
74 rGluTyrIleSerPheIleThrGlyGluAlaAsnGluArgCysGlnArg 91

```

152 erMetSerProAlaAlaMetLeuLysSerArgGlyProValSerGlyAla 168
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
167 ATGAACAGCGTGGCGGTCAGCTCCACTCCACGGT...GCCTCTGGCGCA 213
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
169 AlaMetLeuPro.HisHisHisHisHisAspMetGlnMetHisAlaA 185
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
214 .....CCACCACACACACACACAC.....CACCAGG 242
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
185 laMetTyrGlyThrAlaValProProAlaGlyProHisHis 201
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
243 CGCTC.....GAA 250
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
202 GlyGlyPheLeuMetProHisProGlnGlySerSerHisTyrLeuProTy 218
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
251 CCGCGGATCTGTGACACACATCTCTCGCGCTCG.....CT 288
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
218 rAlaTyrGluProThrTyrGlyGluHisAlaMetAlaAlaTyrTyrG 235
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
289 CGGCTCATGGCGGCGGCGGCGGCGGCGG...GGGCGCGGCGG 335
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
235 lyGlyAlaAlaTyrAlaProGlyAsn..... 243
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
336 GCGGCGCCACAGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 385
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
244 .....GlyGlySerGlyAspGlySerGlyGlyGlyGlyGlyGly 258
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
386 GCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 435
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
258 rAlaSerHisThrProGlnGlySer.....GlyGlyLeuGluHisProH 273
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
436 CGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 485
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
273 isPro 274
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
486 ACCCG 490

```

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-845-998-7

seq_documentation_block:

; Sequence 7, Application US/08845998

; Patent No. 5879892

; GENERAL INFORMATION:

; APPLICANT: Van Baren, Nicolas

; APPLICANT: Coullie, Pierre G.

; APPLICANT: De Smet, Charles

; APPLICANT: Lucas, Sophie

; APPLICANT: Boon, Thierry

; TITLE OF INVENTION: LEUKEMIA ASSOCIATED GENES

; NUMBER OF SEQUENCES: 16

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Wolf, Greenfield & Sacks, P.C.

; STREET: 600 Atlantic Avenue

; CITY: Boston

; STATE: MA

; COUNTRY: US

; ZIP: 02210

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/845,998

; FILING DATE:

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Van Amsterdam, John R.

; REGISTRATION NUMBER: 40,212

; REFERENCE/DOCKET NUMBER: L0461/7008

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (617)720-3500

; TELEFAX: (617)720-2441

; INFORMATION FOR SEQ ID NO: 7:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 4524 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; MOLECULE TYPE: cDNA

; HYPOTHETICAL: NO

; ANTI-SENSE: NO

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 174..1433

; US-08-845-998-7

alignment_scores:

Quality: 167.50 Length: 169

Ratio: 1.971 Gaps: 11

Percent Similarity: 50.296 Percent Identity: 35.503

alignment_block:

US-09-435-054-2 x US-08-845-998-7 ..

Align seg 1/1 to: US-08-845-998-7 from: 1 to: 4524

```

125 ArgGluPheGluGlyAspAlaArgGlyValGly.....Le 136
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
326 CGACACGCTGCTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 375
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
136 uValProGlyAlaAlaProSer...ArgGlyGlyAspHisHisProHis 152
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
376 TGTCACAGGCGAGAGCCATCTTCAAGCGGAGCGGCGGCGGCGGCGG 425
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
152 erMetSerProAlaAlaMetLeuLysSerArgGlyProValSerGlyAla 168
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
426 ATGAACAGCGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 472
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
169 AlaMetLeuPro.HisHisHisHisHisAspMetGlnMetHisAlaA 185
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
473 .....CCACCACACACACACACAC.....CACCAGG 501
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
185 laMetTyrGlyGlyThrAlaValProProAlaGlyProHisHis 201
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
502 CGCTC.....GAA 509
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
202 GlyGlyPheLeuMetProHisProGlnGlySerSerHisTyrLeuProTy 218
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
510 CCGGCGCATCTGTGACACATCTCTCGCGCTCG.....CT 547
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
218 rAlaTyrGluProThrTyrGlyGluHisAlaMetAlaAlaTyrTyrG 235
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
548 CGGCTCATGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 594
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
235 lyGlyAlaAlaTyrAlaProGlyAsn..... 243
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
595 GCGGCGCCACAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 644
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
244 .....GlyGlySerGlyAspGlySerGlyGlyGlyGlyGlyGly 258
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
645 GCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 694
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
258 rAlaSerHisThrProGlnGlySer.....GlyGlyLeuGluHisProH 273
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
695 CGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 744
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
273 isPro 274
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
745 ACCCG 749

```

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-206-537-7

seq_documentation_block:

; Sequence 7, Application US/09206537

alignment_scores:		
Quality:	136.50	Length: 312
Ratio:	0.968	Gaps: 15
Percent Similarity:	45.192	Percent Identity: 23.718

```

alignment_block:
US-09-435-054-2 x US-09-041-886-10
Align seg 1/1 to: US-09-041-886-10 from: 1 to: 3715

8 ProAlaAlaGlyAlaGlnGlySerAlaAlaGlyAlaAsnGln 24
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1257 CCTGGGTGTGGAGGCGTTGGAGCATCTGATCCAGGGAAACAGCTTCGGG 1306
24 yGlyAlaAlaGlnGlnHisAlaAlaProAlaIleArgGluGlnAspArg. 40
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1307 GGGATTG.....CATGTACGCCCCACTTTTGGAGTTCCACCCCGCT 1347
41 .....LeuMetProIleAlaAsnValIleArgIleMetArg 52
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1348 GTGCGTCCCATCTCTGTGCCCATCTGGCCGAATGCAMAGGTTCCTGCT 1397
53 Arg.....ValLeuProAl 57
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1398 AGACGACAGCGCAGGCAAGCAGCTGAAGATACTGCTGAGTATTCCTCTT 1447
57 aHisAlaLysIleSerAspAlaLysGluThrIleGlnGluCysValS 74
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1448 TCAAGGGAGGTACACCAAGGGCTAGAAAGCGCAGAGCCTAGGCTGCTCT 1497
74 erGluTyrIleSerPheIleThrGlyGluAlaAsnGluArgCysGlnArg 90
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1498 GG.....CAGCGCTGCAGCAGG 1514
91 GluGlnArgLysThrIleThrAlaGluAspValLeuTrpAlaMetSerAr 107
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1515 GAGCTCGGGACACTTGAAGTGC..... 1538
107 gLeuGlyPheAspTyrValGluProLeuGlyAlaTyr.LeuHisArg 123
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1539 .....GTCTACCCCTGTCTCTC 1554
124 TyrArgGluPheGluGlyAspAlaArgGlyValGlyLeuValProGlyAl 140
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1555 TACAAGTCC.....GGAGCACTGGACGAGGC 1580
140 aAlaProSerArgGlyAspHisHis..... 149
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1581 AGCTGCTACAGTACGAGTGGGACTACTACAACTTCCACTGGCTCTGGCGG 1630
150 .....ProHisSerMetSerProAlaAlaMet 158
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1631 GACCGCGCGCCCTCCGCGCGCTCCCAT.....CCCCAGCTCGC 1671
159 LeuLysSerArgGlyProVal.....SerGlyAlaAlaMetLeuPr 172
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1672 ATCAAGCTGAGAACCGCTGGACTACGCGAGCGCTGGCGGCTCGGC 1721
172 oHisHisHisHisHisHisAspMetGlnMetHisAlaAlaMetTyrGlyC 189
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1722 GGCAGCTGCGCTATGGGACCTG.....GGAGCCTGCATGGCG 1762
189 lyThrAlaValProProAlaGlyProProHisHisGlyGlyPheLeu 205
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1763 CGGTGCGAGCGGACCGGCTGTGGGTACCCCTCAGCGCGCGCT..... 1806
206 MetProHisProGlnGlySerHisTyrLeuProTyrAla..... 219
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1807 .....TCTCATCTTGCACACTCTCTTCACAGCCGCAAGAGG 1844
220 .....TyrGluProThrTyrGlyGlyGluHisAlaMetAlaTyrT 234
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1845 CCAGTTGTATGGACCGTGTGGTGTGGTGGG..... 1875
234 yrGlyGlyAlaAlaTyrAlaProGlyAsnGlyGlySerGlyAspGlySer 250
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1876 ..GGTGTGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1923
seq_name: /cgn2_6/ptodata/2/Ina/5A_COMB.seq:US-08-073-384C-2
seq_documentation_block:
; Sequence 2, Application US/08073384C
; Patent No. 5541311
; GENERAL INFORMATION:
; APPLICANT: Dahlberg, James E.
; APPLICANT: Lyamichev, Victor I.
; APPLICANT: Brow, Mary Ann D.
; TITLE OF INVENTION: SYNTHESIS-DEFICIENT THERMOSTABLE DNA
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HAVERSTOCK, MEDLEN & CARROLL
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/073,384C
; FILING DATE: 04-JUN-1993
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/986,330
; FILING DATE: 07-DEC-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Carroll, Peter G.
; REGISTRATION NUMBER: 32,837
; REFERENCE/DOCKET NUMBER: FORS-00613
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/705-8410
; TELEFAX: 415/397-8338
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2496 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-073-384C-2
alignment_scores:
Quality: 134.50 Length: 317
Ratio: 0.934 Gaps: 17
Percent Similarity: 45.426 Percent Identity: 26.498
alignment_block:
US-09-435-054-2 x US-08-073-384C-2
Align seg 1/1 to: US-08-073-384C-2 from: 1 to: 2496
19 GlyGlyAlaAsnGlyGlyAlaAlaGln.GlnHisAlaAlaProAlaI 35
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1035 GGGGTGCGGGGAATCCTGGC.....CAAGGACCTGGCGGTTTGGGCC 1078
35 leArgGluGlnAspArgLeuMetProIleAlaAsnValIleArgIleMet 51
|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```
1079 TCGGGAGGGCTGACCTCTTCCAGAGGACGACCCCATGCTCCTG... 1125
52 ArgArgValLeuProAlaHisAlaLysIleSerAspAlaLysGluTh 68
1126 .....GCTACCTCTGACCCCTCCACACAC 1154
68 rleGlnGluCysValSerGluTyrIleSerPheIleThrGlyAla 85
1155 CCCTGAGGGGTGCGCGGTACGGGGGAGTGCAGGAGGATGCGG 1204
85 snGluArgCysGlnArgGluGlnArgLysThrIleThrAlaGluAspVal 101
1205 GGGAGAGGGCCCTCTGCGCCGAGCGCTCTCCAGACCCCTAAGAGCGC 1254
102 LeuTrpAlaMetSerArgLeu.....GlyPheAspAspTyrValGluPr 116
1255 CTTAAGGGAGAAACCCCTGCTTTCAGGAGGCTGAGAGGCTGAGAGCC 1304
116 oLeuGlyAlaTyrLeuHisArgTyrArgGluPheGluGlyAspAlaArg 133
1305 GCTTTCGGGTGTGGCCCGATGAGGC....CACGGGGTCCGGCTG 1350
133 lyValGlyLeuValProGlyAlaAlaProSerArgGlyGlyAspHis 149
1351 GAGTGGCTTACCTCCAGGCCCTCTCCCTGGAGGTGAGGCGGAGTGGC 1400
150 ProHisSerMetSerProAlaAlaMetLeuLysSerArgGlyProVal.. 165
1401 CCA.....GCTGAGGAGGAGGTCTTCCGCCGTGGCCGCCACCCCTCA 1444
166 .....SerGlyAlaAlaMetLeu..... 171
1445 ACCTCAACTCCCGGACCACTGAGCGGTCTCTTTCAGGAGTGGGC 1494
172 .....ProHisHisHisHisHisAspMetGln...MethHisAlaAla 185
1495 CTGCTGCCATCGCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1544
186 MetTyrGlyThrAlaValProProAlaGlyProProHisHisGlu 202
1545 TGGCGTCTGGA.....GCCCTGGAGAGGAGGAGGAGGAGGAGGAGG 1585
202 yGlyPheLeuMetProHisPro.....GlnGlySerS 213
1586 A.....CCGATCTCTGACCTACGCGGAGGCTCACCAGCTCAAG 1623
213 erHisTyr..... 215
1624 AACACCTACATAGACCCCTGCGCGCTGCTCCACCCCAAGAGCGCGC 1673
216 LeuProTyrAlaTyrGluProThrTyrGlyGlyGluHisAlaMetAla.. 231
1674 GCTCCACACCCGCTTCAACACAG...GCCACCCGAGGAGGAGGAGG 1720
232 .....AlaTyrTyrGlyGlyA 237
1721 CCAGCTCCGACCCCAACCTGCAGAACATCCCGTGGCACCCTCTGGGC 1770
237 laAlaTyrAlaPro.....GlyAsnGlyGlySerGlyAspGly... 249
1771 CAGCGCATCCGCGAGCTTCTGTCGCGAGGAGGAGGAGGAGGAGGAGG 1820
250 .....SerGlySerGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 258
1821 CTGGAGTACACCCAGATGACCTGCGGTCTGCGCCACCTCTCCGGGG 1870
258 rLaSerHisThrProGlnGlySerGlyGlyLeuGluHisProHisPro 274
1871 ACGAGAACCTGATCGGGTCTTTCAGGAGGAGGAGGAGGAGGAGGAGG 1919
seq_name: /cgn2_6/ptdata/2/ina/5A_COMB.seq.us-08-254-359A-2
seq_documentation_block:
```

```
; Sequence 2, Application US/08254359A
; Patent No. 5614402
; GENERAL INFORMATION:
; APPLICANT: DAHLBERG, JAMES E.
; APPLICANT: LYAMICHEV, VICTOR I.
; APPLICANT: BROW, MARY ANN D.
; TITLE OF INVENTION: 5' NUCLEASES DERIVED FROM THERMOSTABLE
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HAVERSTOCK, MEDLEN & CARROLL
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/254,359A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/073,384
; FILING DATE: 06-JUN-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/986,330
; FILING DATE: 07-DEC-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: CARROLL, PETER G.
; REGISTRATION NUMBER: 32,837
; REFERENCE/DOCKET NUMBER: FORS-01000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2496 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-254-359A-2

alignment_scores:
    Quality: 134.50      Length: 317
    Ratio: 0.934        Gaps: 17
    Percent Similarity: 45.426    Percent Identity: 26.498

alignment_block:
US-09-435-054-2 x US-08-254-359A-2 ..
Align seg 1/1 to: US-08-254-359A-2 from: 1 to: 2496
19 GlyGlyAlaAsnAsnGlyGlyAlaAlaGln.GlnHisAlaAlaProAlaI 35
1035 GGGGGTGGGGGATCTGCG.....CAAGGACCTGGCGGTTTGGCCC 1078
35 leArgGluGlnAspArgLeuMetProIleAlaAsnValIleArgIleMet 51
1079 TCGGGAGGAGGCTGACCTCTTCCAGAGGACGACCCCATGCTCCTG... 1125
52 ArgArgValLeuProAlaHisAlaLysIleSerAspAlaLysGluTh 68
1126 .....GCCCTCTTCTGACCCCTCCACACAC 1154
68 rleGlnGluCysValSerGluTyrIleSerPheIleThrGlyAlaAla 85
1155 CCCTGAGGGGTGCGCGGTACGGGGGAGTGCAGGAGGATGCGG 1204
```



```

85  snGluArgCysGlnArgGluGlnArgLysThrIleThrAlaGluAspVal 101
1205  GGGAGAGGGCCCTCCCTGGCGCAGCGCCTCTTCCAGACCCCTAAAGGAGCGC 1254
102  LeuTrpAlaMetSerArgLeu.....GlyPheAspAspTyrValGluPr 116
1255  CTTAAGGGAGAGACGCCTCTGTGGCTTTTACGAGGAGGTGGAGAAGCC 1304
116  oLeuGlyAlaTyLeuHisArgTyrArgGluPheGluGluGlyAspAlaArg 133
1305  GCTTTCCCGGGTGTGGCCCGATGGAGC....CACGGGGTCCGGCTG 1350
133  lyValGlyLeuValProGlyAlaAlaProSerArgGlyGlyAspHisHis 149
1351  GACGTGGCCCTACTCAGGCGCCTCTCCCTGGAGGTGGAGCGGAGTGGC 1400
150  ProHisSerMetSerProAlaAlaMetLeuLysSerArgGlyProVal... 165
1401  CCA.....GCTGGAGGAGGAGGTCTCCGCTGGCCGCCACCCCTTCA 1444
166  .....SerGlyAlaAlaMetLeu..... 171
1445  ACCTCAACTCCGCGACAGCTGGAGCGGGTCTCTTTGACGAGCTGGGC 1494
172  ....ProHisHisHisHisAspMetGln...MethisAlaAla 185
1495  CTGCCTGCCATCGCAACGAGGAGAGAGCGGGAACGCTCCACCAAGCGC 1544
186  MetTyrGlyGlyThrAlaValProProAlaGlyProHisHisG1 202
1545  TGCCGTGCTGA.....GGCCCTGCGAGAGGCCACCCCATCGTGG 1585
202  yGlyPheLeuMetProHisPro.....GlnGlySerS 213
1586  A.....CCGATCTCGAGTACCGGGAGCTCACCAAGCTCAAG 1623
213  erHisTyr..... 215
1624  AACACCTACATAGACCCCTGCCCGCCCTGTCTCCACCCCAAGACCGCGC 1673
216  LeuProTyrAlaTyGluProThrTyrGlyGlyGluHisAlaMetAla.. 231
1674  GCTCCACACCGCTTCAACAGAC...GGCCAGCCGACCGGCGAGGCTTT 1720
232  .....AlaTyrTyrGlyGlyA 237
1721  CCAGCTCCGACCCCAACCTGCAGAAATCCCGCGTGGCAGCCCTCTGGGC 1770
237  laAlaTyAlaPro.....GlyAsnGlyGlySerGlyAspGly... 249
1771  CACGCGATCCGCGAGCTTCTGTGGCCGAGGAGGGCTGGTGTCTGGTGGT 1820
250  .....SerGlySerGlyGlyGlyGlyGlyse 258
1821  CTTGGACTACGACGAGATTGAGTCTCGGTCTGGCCCACTCTCCGGGG 1870
258  rAlaSerHisThrProGlnGlySerGlyGlyLeuGluHisProHisPro 274
1871  ACAGAAACCTGATCCGGGTCTTTTCAGAGGGGAGGACATCTCACACCCA 1919

```

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-483-043-2

```

seq_documentation_block:
; Sequence 2, Application US/08483043
; Patent No. 5691142
; GENERAL INFORMATION:
; APPLICANT: Dahlberg, James E.
; APPLICANT: Lyemichnev, Victor I.
; APPLICANT: Brow, Mary Ann D.
; TITLE OF INVENTION: SYNTHESIS-DEFICIENT THERMOSTABLE DNA
; TITLE OF INVENTION: POLYMERASE
; NUMBER OF SEQUENCES: 29

```

CORRESPONDENCE ADDRESS:
 ADDRESSEE: HAVERSTOCK, MEDLEN & CARROLL
 STREET: 220 Montgomery Street, Suite 2200
 CITY: San Francisco
 STATE: California
 COUNTRY: United States of America
 ZIP: 94104
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent In Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/483,043
 FILING DATE: 06-JUN-1995
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/073,384
 FILING DATE: 04-JUN-1993
 APPLICATION NUMBER: US 07/986,330
 FILING DATE: 07-DEC-1992
 ATTORNEY/AGENT INFORMATION:
 NAME: Cartoll, Peter G.
 REGISTRATION NUMBER: 32,837
 REFERENCE/DOCKET NUMBER: FORS-00613
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 415/705-8410
 TELEFAX: 415/397-8338
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2496 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 US-08-483-043-2

```

alignment_scores:
  Quality: 134.50      Length: 317
  Ratio: 0.934        Gaps: 17
  Percent Similarity: 45.426  Percent Identity: 26.498

alignment_block:
  US-09-435-054-2 x US-08-483-043-2  ..

Align seg 1/1 to: US-08-483-043-2 from: 1 to: 2496

19  GlyGlyAlaAsnAsnGlyGlyAlaAlaGln.GlnHisAlaAlaProAlaI 35
1035 GGGGGTGGGGGAAATCCCTGGC.....CAAGGACCTCGGGTTTGGCCC 1078
35  LeArgGluGlnAspArgLeuMetProIleAlaAsnValIleArgIleMet 51
1079 TGGGGAGGGCGCTGCACCTCTTCCCAGAGGACGCCCATGTCTCTG... 1125
52  ArgArgValLeuProAlaHisAlaLysIleSerAspAlaLysGluThr 68
1126 .....GCCTACCTCTCGGACCCCTCCACACCAC 1154
68  rIleGlnGluCysValSerGluTyrIleSerPheIleThrGlyGluAlaA 85
1155 CCCTGAGGGGTGGCCCGCGCTTACGGGGAGTGGCAGGAGTGGCG 1204
85  snGluArgCysGlnArgGluGlnArgLysThrIleThrAlaGluAspVal 101
1205 GGGAGAGGGCCCTCTGCGCGAGCGCCTCTTCCAGACCCCTAAAGAGCGC 1254
102  LeuTrpAlaMetSerArgLeu.....GlyPheAspAspTyrTyrValGluPr 116
1255 CTTAAGGGAAGAAGAACGCTCTCTTGGCTTTACGAGGAGGTGGAGAAGCC 1304
116  oLeuGlyAlaTyrLeuHisArgTyrArgGluPheGluGlyAspAlaArgG 133

```

```

1305 GCTTCCGGGTGTTGGCCGGATGAGGC.....CACGGGGTCCGGCTG 1350
133 lyValGlyLeuValProGlyAlaAlaProSerArgGlyGlyAspHisHis 149
1351 GACGTGGCCTACCTCCAGGCGCTCTCCCTGGAGGTGGAGGAGGTGGC 1400
150 ProHisSerMetSerProAlaAlaMetLeuLysSerArgGlyProVal.. 165
1401 CCA.....GCTGGAGGAGGAGGTCTTCCGCTGGCGGCCACCCCTTCA 1444
166 .....SerGlyAlaAlaMetLeu..... 171
1445 ACCTCACTCCCGGACACGCTGGAGCGGTGCTCTTTGACGAGCTGGGC 1494
172 .....ProHisHisHisHisHisMetGln...MetHisAlaAla 185
1495 CTGCCTGCCATCGCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1544
186 MetTyrGlyGlyThrAlaValProProAlaGlyProProHisHisG1 202
1545 TCCCGTGTGGA.....GGCCCTGGAGAGGCCACCCCTCGTGG 1585
202 yGlyPheLeuMetProHisPro.....GlnGlySers 213
1586 A.....CCGCATCTGTCAGTACCGGAGGAGCTCACCAAGCTCAAG 1623
213 erHisTyr..... 215
1624 AACACCTACATAGACCCCTCCCGCCCTGGTCCACCCCAAGACCGGCG 1673
216 LeuProTyrAlaTyrGluProThrTyrGlyGlyGluHisAlaMetAla.. 231
1674 GCTCCACACCCGCTTCAACACAGAC...GGCCACCGCCACGGGAGGCTT 1720
232 .....AlaTyrTyrGlyGlyA 237
1721 CCAGCTCCGACGCCAACCTGCAGAACATCCCGCTGGCGACCCCTCTGG 1770
237 laAlaTyrAlaPro.....GlyAsnGlyGlySerGlyAspGly... 249
1771 CAGCGCATCCGCGGAGCCTTCGTGGCGGAGGAGGAGGAGGAGGAGG 1820
250 .....SerGlySerGlyGlyGlyGly 258
1821 CTTGGACTACACCCAGATTGAGCTTCGGGTCTCGGCCACCTCTCCGGG 1870
258 rAlaSerHisThrProGlnGlySerGlyGlyLeuGluHisProHisPro 274
1871 ACAGAACTGATCCGGGTCTTTCAGGAGGAGGAGGAGGAGGAGGAGG 1919

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:us-08-481-238-2

seq_documentation_block:
; Sequence 2, Application US/08481238
; Patent No. 5795763
; GENERAL INFORMATION:
; APPLICANT: DAHLBERG, JAMES E.
; APPLICANT: LYAMICHEV, VICTOR I.
; APPLICANT: BROW, MARY ANN D.
; TITLE OF INVENTION: SYNTHESIS-DEFICIENT THERMOSTABLE DNA
; NUMBER OF INVENTION: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MEDLEN & CARROLL
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

```

172ProHisHisHisHisHisAspMetGln....MethisAlaAla 185
|||:||||| :||| :|||
1495 CTGCGTCGCATCGGCAAGACGAGAGAGAGCGGGAAACGCTCCACCAGCGC 1544
186 MetTyrGlyGlyThrAlaValProProAlaGlyProProHisHisG1 202
:||||| :||| :|||||:|||||:|||||
1545 TGCCGTGCTGGA.....GGCCCTGGGAGAGGCCACCCCATCGTGG 1585
202 yGlyPheLeuMetProHisPro.....GlnGlySers 213
| :||||| :|||
1586 A.....CGGCATCTCTGCAGTACCGGGAGCTCACCAAGCTCAAG 1623
213 erHistyr..... 215
:|||||
1624 AACCTACATAGACCCCTGCCGCCCTGTCTCCACCCCAAGACCGGCG 1673
216 LeuProTyrAlaTyrGluProThrTyrGlyGlyGluHisAlaMetAla.. 231
|||:| :||| :||| :|||||:|||||
1674 GCTCCACACCCGCTTCAACCGAC...GGCCACCGCCACGGGCGGCTTT 1720
232AlaTyrTyrGlyGlyA 237
|||:| :|||
1721 CCAGCTCCGACCCCAACCTGCAGAACATCCCGCTGCGCACCCCTCTGGGC 1770
237 laAlaTyrAlaPro.....GlyAsnGlyGlySerGlyAspGly... 249
|||:| :||| :|||||:||||| :|||
1771 CAGCGCATCCGCCGAGCCTTCGTGCCCGGAGGAGGCTGGGTGCTGGT 1820
250SerGlySerGlyGlyGlyGlyGly 258
||||| :|||
1821 CTTGGACTACAGCCAGATTGAGCTTCGGGTCTCTGCCCCACCTCTCCGGGG 1870
258 rAlaSerHisThrProGlnGlySerGlyGlyLeuGluHisProHisPro 274
: :||| :||||| :|||||
1871 ACGAAGACCTGATCCGGGTCTTTTCAGGAGGGGAGGACATCCACACCCA 1919

